Computationally Efficient Algorithms and Implementations of Adaptive Deep Brain Stimulation Systems for Parkinson's Disease

Ameer Mohammed

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Supervisors:

Principal: Prof. Andreas Demosthenous

Subsidiary: Prof. Richard Bayford

January 2018

Analogue and Biomedical Electronics Group

Department of Electronic and Electrical Engineering

University College London

Declaration

I, Ameer Mohammed, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been appropriately indicated in the thesis.

To people worldwide suffering from various neurodegenerative diseases.

Abstract

Clinical deep brain stimulation (DBS) is a tool used to mitigate pharmacologically intractable neurodegenerative diseases such as Parkinson's disease (PD), tremor and dystonia. Present implementations of DBS use continuous, high frequency voltage or current pulses so as to mitigate PD. This results in some limitations, among which there is stimulation induced side effects and shortening of pacemaker battery life. Adaptive DBS (aDBS) can be used to overcome a number of these limitations. Adaptive DBS is intended to deliver stimulation precisely only when needed. This thesis presents work undertaken to investigate, propose and develop novel algorithms and implementations of systems for adapting DBS. This thesis proposes four system implementations that could facilitate DBS adaptation either in the form of closed-loop DBS or spatial adaptation. The first method involved the use of dynamic detection to track changes in local field potentials (LFP) which can be indicative of PD symptoms. The work on dynamic detection included the synthesis of validation dataset using mainly autoregressive moving average (ARMA) models to enable the evaluation of a subset of PD detection algorithms for accuracy and complexity trade-offs. The subset of algorithms consisted of feature extraction (FE), dimensionality reduction (DR) and dynamic pattern classification stages. The combination with the best trade-off in terms of accuracy and complexity consisted of discrete wavelet transform (DWT) for FE, maximum ratio method (MRM) for DR and k-nearest neighbours (k-NN) for classification. The MRM is a novel DR method inspired by Fisher's separability criterion. The best combination achieved accuracy measures: F1-score of 97.9%, choice probability of 99.86% and classification accuracy of 99.29%. Regarding complexity, it had an estimated microchip area of 0.84 mm^2 for estimates in 90 nm CMOS process. The second implementation developed the first known PD detection and monitoring processor. This was achieved using complementary detection, which presents a hardware-efficient method of implementing a PD detection processor for monitoring PD progression in Parkinsonian patients. Complementary detection is achieved by using a combination of weak classifiers to produce a classifier with a higher consistency and confidence level than the individual classifiers in the configuration. The PD detection processor using the same processing stages as the first implementation was validated on an FPGA platform. By mapping the implemented design on a 45 nm CMOS process, the most optimal implementation achieved a dynamic power per channel of 2.26 μW and an area per channel of 0.2384 mm^2. It also achieved mean accuracy measures: Mathews correlation coefficient (MCC) of 0.6162, an F1-score of 91.38%, and mean classification accuracy of 91.91%. The third

implementation proposed a framework for adapting DBS based on a *critic-actor* control approach. This models the relationship between a trained clinician (critic) and a neuro-modulation system (actor) for modulating DBS. The critic was implemented and validated using machine learning models, and the actor was implemented using a fuzzy controller. Therapy is modulated based on state estimates obtained through the machine learning models. PD suppression was achieved in seven out of nine test cases. The final implementation introduces spatial adaptation for aDBS. Spatial adaptation adjusts to variation in lead position and/or stimulation focus, as poor stimulation focus has been reported to affect therapeutic benefits of DBS. The implementation proposes dynamic current steering systems as a power-efficient implementation for multi-polar multisite current steering, with a particular focus on the output stage of the dynamic current steering system. The output stage uses dynamic current sources in implementing push-pull current sources that are interfaced to 16 electrodes so as to enable current steering. The performance of the output stage was demonstrated using a supply of 3.3 V to drive biphasic current pulses of up to 0.5 mA through its electrodes. The preliminary design of the circuit was implemented in 0.18 μm CMOS technology.

Acknowledgement

I would like to thank my primary supervisor, Prof. Andreas Demosthenous for guiding me throughout my PhD. It was really a good journey, and I so much appreciate and cherish his mentorship over these years. His invaluable and tireless support and contributions in my work are very much appreciated. Working with him has really been an honour, as I have gained a lot academically, professionally and morally. I would like to acknowledge the efforts of my second supervisor, Prof. Richard Bayford, for his immense support so far. I have always come out better informed after every discussion with him. In addition, I would like to thank Prof. Peter Brown of the Division of Clinical Neurology at the University of Oxford for providing the test datasets.

I am also thankful to my colleagues in office 1006, who have really been helpful and have been like a second family to me over these years. Also, I would like to acknowledge my sponsors, the Presidential Special Scholarship for Innovation and Development (PRESSID) Nigeria, who have provided this platform for my development.

Finally, a special and very big thank you goes to my parents, my wife, my daughter, my siblings, relatives and my friends for their constant love, encouragement and understanding throughout my life.

Contents

Abbreviations

ACF	Autocorrelation Function
AD	Alzheimer's Disease
aDBS	Adaptive Deep Brain Stimulation
ADC	Analog-to-Digital Converter
ADF	Augmented Dickey-Fuller
ADHD	Attention Deficit Hyperactivity Disorder
AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARV	Average Rectified Value
ASIC	Application Specific Integrated Circuit
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
BMI	Brain-Machine-Interface
BOLD	Blood-Oxygen-Level-Dependent
BSS	Blind Source Separation
CCS	Current Controlled Stimulation
CFM	Carbon Fibre Microelectrodes
ChgCS	Charge Controlled Stimulation
CP	Choice Probability
DBS	Deep Brain Stimulation
DCS	Dynamic Current Sources
dMRI	Diffusion Magnetic Resonance Imaging
DR	Dimensionality Reduction
DWT	Discrete Wavelet Transform
ECoG	Electrocorticograhy
EDP	Energy-Delay Product
EEG	Electroencephalography
ERP	Event-Related Potential
ET	Essential Tremor
ETI	Electrode-Tissue Interface
FDA	Food and Drug Administration
FE	Feature Extraction
FFT	Fast Fourier Transform
FIFO	First-in First-out
fMRI	Functional Magnetic Resonance Imaging
FPGA	Field Programmable Gate Array
FSCV	Fast Scan Cyclic Voltammetry
FSM	Finite State Machine
GMM	Gaussian Mixture Model

Gpe	Globus Pallidus Externa
Gpi	Globus Pallidus pars Internus
HDL	Hardware Description Language
HMM	Hidden Markov Model
HPF	High-Pass Filter
ICA	Independent Component Analysis
IPG	Implantable/Internal Pulse Generator
k-NN	k-Nearest Neighbours
LDA	Linear Discriminant Analysis
L-dopa	Levodopa
LFP	Local Field Potential
LNA	Low Noise Amplifier
LPF	Low Pass Filter
LR	Logistic Regression
LUT	Look up Tables
M1	Primary Motor Cortex
MA	Moving Average
MCC	Mathews Correlation Coefficient
MIMO	Multiple Input Multiple Output
MINCS	Mayo Investigational Neuromodulation Control System
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MRCP	Movement-Related Cortical Potentials
MRM	Maximum Ratio Method
NHS	National Health Service
NIRS	Near Infrared Spectroscopy
NOP	Number of Operations
OCD	Obsessive-Compulsive Disorder
PACF	Partial Autocorrelation Function
PD	Parkinson's Disease
PDF	Probability Density Function
PDP	Power-delay product
PID	Proportion-Integral-Derivative
PSD	Power Spectral Density
RAM	Random Access Memory
RBF	Radial Basis Function
RNS	Responsive Neurostimulator System
ROC	Receiver Operating Characteristic
RTL	Register Transfer Level
SCI	Spinal Cord Injury
SN	Substantia Nigra
SNR	Signal to Noise Ratio
SoC	System on Chip
SR	Sensorimotor Rhythms
SRAM	Static Random Access Memory
St	Striatum
STFT	Short Time Fourier transform
STN	Subthalamic Nucleus
SVM	Support Vector Machine
SZ	Schizophrenia
TS	Tourette's Syndrome
UART	Universal Asynchronous Receiver/Transmitter

UPDRS	Unifed Parkinson's Disease Rating Scale
Vc	Ventral Caudal
VCR	Voltage Controlled Resistor
VCS	Voltage Controlled Stimulation
VIM	Ventral Intermediate Nucleus
WCE	Weighted Classification Error
WSS	Weak/Wide Sense Stationary

List of Figures

List of Tables

Chapter 1

Introduction

The field of neurostimulation dates back to the 18th century when the Italian physician and physicist Luigi Galvani discovered what he called "animal electricity" when he activated skeletal muscle contraction with an electric current. Since that time there have been significant advances in the understanding of basic anatomy and physiology of body tissue when in contact with electricity. Cortical stimulation dates back to the 19th century when the first direct electrical stimulation of the cortex was recorded [1]. Even though cortical stimulation has existed from the 19th century, therapeutic Deep brain stimulation (DBS) was discovered by Benabid et al. [11]. DBS is the conventional therapy for treating various neurological disorders such as Parkinsons disease (PD), essential tremor (ET) and dystonia. From then on, research in DBS has witnessed a remarkable surge in interest. A particularly significant milestone was the approval of DBS as a therapy for PD in 1997 by the Food and Drug Administration (FDA). At present, there is a growing interest in using DBS as a therapy for neuropsychiatric disorders like medication resistant cases of depression, obsessive-compulsive disorder (OCD) and Tourette's syndrome (TS) [12]. Essentially, this chapter provides a context for clinical DBS for PD and some of its challenges that motivated this work. The sections of this chapter are organised as follows: Section 1.1 provides a brief introduction on therapeutic interventions for mitigating PD notably, clinical DBS and its limitations. Section 1.2 provides the objectives of the work as well as the research gaps targeted by this work. Section 1.3 lists the publications that have so far resulted from this work. And finally, Section 1.4 summarises the major points addressed in subsequent chapters of the thesis.

1.1 Clinical DBS

In the UK, based on figures from the National Health Service (NHS), it is estimated that 127,000 people are suffering from PD [13], these numbers are on the rise. PD is one of the most common neurodegenerative diseases leading to impairment of both motor and non-motor functions [14]. It manifests mostly in elderly persons but can occasionally be found in younger patients. Pathologically, PD is caused by degeneration of the dopamine-producing neurons in the substantia nigra (SN), this starves the brain of dopamine [15]. Thus, affecting basic tasks in patients since dopamine is a neu-

rotransmitter in the brain that assists in functions like movement, memory, cognition among others. Symptoms manifest by the presence of rest tremor, postural instability, rigidity and bradykinesia [16].

At its advanced stage, therapeutic interventions like ablative surgery and clinical DBS are administered. The surgical sites for these interventions are the ventral intermediate nucleus (VIM) of the thalamus, the globus pallidus pars internus (GPi), or the subthalamic nucleus (STN), to restore motor functions. The STN has been identified as the most effective site for mitigating PD symptoms [17]. PD treatment using ablative surgery results in serious side effects by exacerbating speech, gait, and cognitive disorders; and may even cause hemiballismus [18]. This makes it less advantageous compared to DBS, which has reversible effects. For clinical DBS, the choice of stimulation site depends on the dominant symptom of each patient. Tremor dominant PD is mitigated by stimulating the VIM, while bradykinesia and rigidity dominant PD can be reduced by stimulating the GPi or STN. Furthermore, DBS is most effective when used in a bilateral configuration. Figure 1.1 illustrates DBS being unilaterally administered at the STN. Currently, PD is the neurological condition most widely treated by DBS [1]. Though the underlying mechanism of DBS therapy is still unclear [1, 19, 20]. Nonetheless, it improves the quality of life of patients, even though it's very expensive in terms of life long maintenance. Its high cost is as a result of the incurable nature of PD [3]. Continuous repetitive stimulation (mostly at 130 Hz) causes some long term effects which result in impaired cognitive abilities and motor functions. It affects coordinated neuronal communication necessary for proper cognitive functions [21]. Presently, post-surgery programming of stimulation parameters by trained clinicians can take up to a year or more [22]. This is because stimulation parameters are adjusted heuristically by using visual symptoms as the only feedback, after which follow up visits by patients are required in order to adjust stimulation to changes in patient condition.

In DBS, stimulation parameter setting is a very tedious process as every patient has a unique set of stimulation parameters. Figure 1.1 illustrates an implantable/internal pulse generator (IPG) for DBS, showing the unilateral stimulating leads with cylindrical electrodes at the tip of each lead [23]. In chronic deep brain stimulation, these leads can migrate which may degrade the effectiveness of the stimulation, requiring the need to adjust stimulation to these changes constantly. More so, due to the progressive nature of PD, selected parameters may become obsolete with time, necessitating the need to track patient pathophysiology. Both of these can be adjusted for using adaptive deep brain stimulation (aDBS) [24]. There are two forms of aDBS: closed-loop adaptation and spatial adaptation. Closed-loop adaptation is ideally designed to track changes in patient pathophysiology and correspondingly adjust stimulation. This is necessary because continuous stimulation is suggested to result in side effects and shortening of pacemaker battery life [25]. On the other hand, spatial adaptation adjusts to variation in lead position and/or stimulation focus, as poor stimulation focus has been reported to affect therapeutic benefits [26].

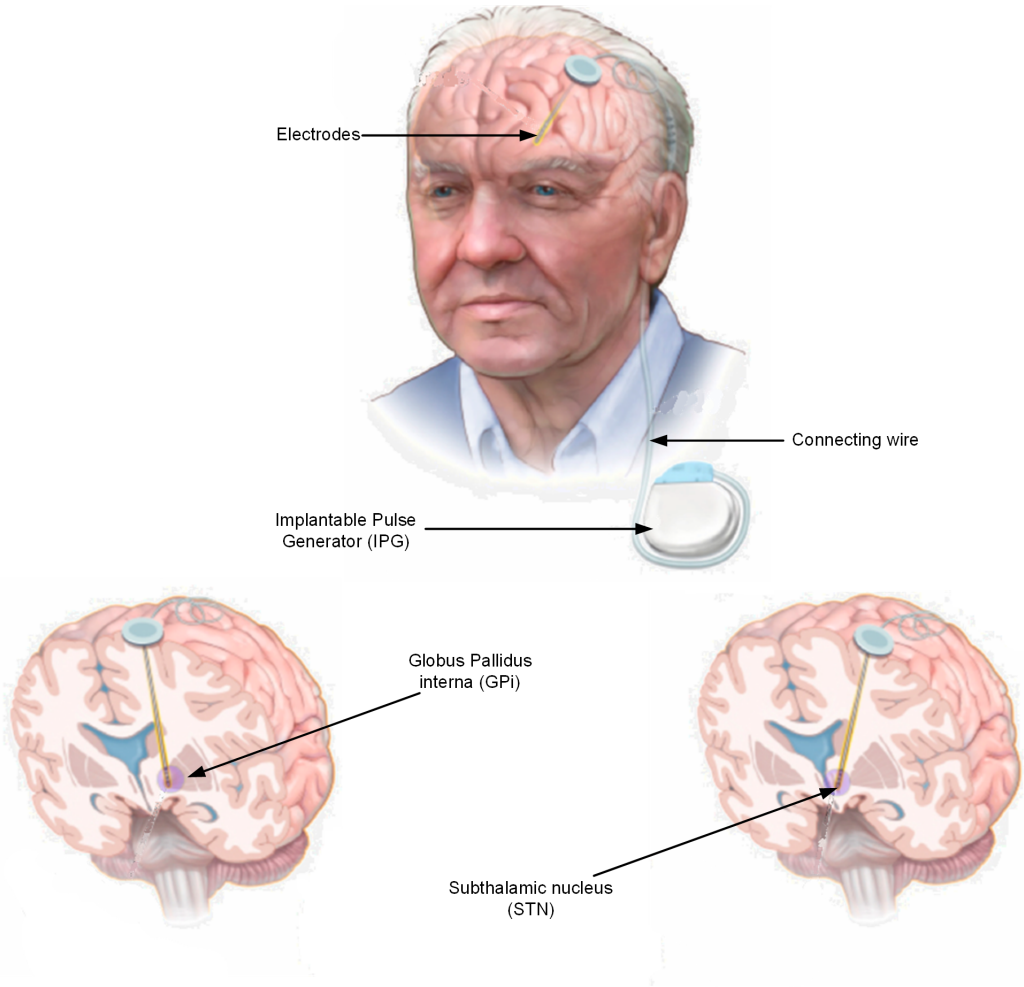


Figure 1.1: Components of a DBS implant and major stimulation sites for clinical DBS (adapted from [1]).

1.1.1 Theories on How DBS Works

Clinical DBS has emerged as an effective therapy for PD and other movement disorders, yet its operating mechanism is still intensely under debate. The choice of the stimulation parameters (i.e. pulse width, frequency and amplitude) used has been reported to affects the effectiveness of DBS in patients [27]. What is not clear is how different stimulation patterns and parameters influence the efficacy of DBS. There are various theories on how DBS induces beneficial effects in patients. The main mechanisms posited for DBS are "inhibition", "excitation" and "disruption".

The *inhibition hypothesis* is the oldest. It was suggested after DBS was found to have the same beneficial effect on PD symptoms as lesion therapy [28, 29]. STN-DBS and GPi-DBS were believed to inhibit local neuronal activity, by reducing the firing rate of surrounding neurons. This was corroborated by recordings obtained around the stimulation sites of STN-DBS in PD patients and non-human primate models [30, 31]. Using the firing rate model in movement disorders which claims that parkinsonian impairments result in abnormally increased firing pattern in the STN and/or GPi, and DBS mitigates the motor symptoms by reducing this increased firing. However,

others point out that, how then has it been possible to treat motor symptoms in situations in which the GPi shows low activity [32]. On the contrary, other studies suggest that the most natural explanation for DBS is the *excitation hypothesis*, where stimulation depolarises neuronal elements. They believe DBS activates the STN (or GPi) output, thus jamming abnormal pathological activity in basal ganglia circuits resulting from motor impairments [33]. In the disruption hypothesis [32], DBS exerts therapeutic influence by dissociating input and output signals, thereby disrupting the flow of abnormal information to the stimulation site. In this theory, therapeutic DBS is allegedly believed to disconnect the abnormal coupling in the basal ganglia as a result of PD.

An emerging theory is the *nonexclusive hypothesis* [34]. Since the other theories present exclusive mechanisms for DBS, the nonexclusive theory suggests that DBS is a result of many (nonexclusive) mechanisms including modulation of oscillatory activity, neurogenesis, synaptic plasticity, neurochemical effects and neuroprotection. The mechanisms vary in importance and manifestation depending on the condition being treated and the brain structure (target) being stimulated. At present, the mechanisms of DBS are still under debate and a proper understanding of this mechanisms could lead to improved therapeutic interventions.

1.1.2 Limitations of Clinical DBS

PD Patients are treated with a combination of medication and therapeutic stimulation; however, therapeutic interventions are ineffective in 30% of cases. Therapeutic stimulation is rendered ineffective by disease progression, environmental factors, mechanical factors, and behaviourally induced changes in network activity. As a result, additional sessions to manually adjust stimulation settings may be required [35]. The procedure can be very costly and time-consuming. This is because only a fraction of the stimulation parameter space can be practically explored during each session. Moreover, as PD progresses the dominant symptoms may change which interferes with DBS device programming and there may need to modify stimulation settings. Programming is dependent on the stimulated target, electrode orientation relative to the target, the disorder treated, and the dominant symptoms for the given disorder. Other shortcomings of clinical DBS include stimulation induced side effects for example drooling, flushing dysarthria, and ocular deviation [26]. They are mainly caused by continuous stimulation as well as stimulation field spread beyond target areas [24]. Stimulation induced side effects can be triggered mainly due to three scenarios, namely:

- False-positive detection results in administering stimulation when it is not required, and this may lead to stimulation induced side effects [36].
- False-negative detection may result in the non-administering of stimulation when it is required, which may worsen patient condition [37].
- An additional concern is the drift of electrodes in chronic DBS. Based on current electrode designs, DBS electrodes are cylindrical around the lead. The change in

electrode position as is normally the case in chronic DBS, inhibits access to target stimulation sites. The requirements for precision are high as minimal variation in DBS lead position can cause a large variation in therapeutic outcome. Increasing stimulation intensity to overcome these may be counterproductive as unintended targets could be touched which may result in stimulation-induced side effects.

The aforementioned limitations can be resolved using aDBS [24, 38]. For closed-loop adaptation, stimulation is optimised by controlling the stimulation intensity, timing and direction using feedback signals from the stimulation site. For adaptation to changes in lead position, directional steering, which enables control over stimulation focus can be used. Adaptive DBS has the potential to improve efficacy, reduce power consumption and reduce side effects. The whole point of aDBS is to implement control techniques that can adjust stimulation parameters in real-time according to quantifiable and objective neurochemical, physiological, and behavioural changes while reducing the frequency of clinical interventions. Such that clinical benefits are optimised while side effects are reduced; and as a secondary effect battery depletion is also minimised [24, 25, 38]. This is necessary because implantable DBS systems have certain resource constraints, for example, the power density of the surrounding chip has to be much less than the power density for brain tissue damage, which is $800\,\mu W/mm^2$ [39]. It is against this backdrop that the major research objectives are identified.

1.2 Objectives

To enable the development of fully implantable systems for adaptive DBS, research gaps around hardware efficiency in implementations for aDBS were identified and explored. And in no particular order of importance below is a list of the major research gaps targeted:

- With hardware implementation in mind, there is a need to identify, quantify and exhaustively evaluate a combination of algorithms that can be used for PD detection based on their complexity and accuracy trade-off. Two other contributions resulted from pursuing this objective: the first involved using statistical methods to generate semi-synthetic local field potentials (LFP) datasets, which was the first of its kind using real LFP recordings obtained from PD patients. Semi-synthetic generation of test datasets was necessary due to the paucity of validation datasets. The other contribution presented dynamic detection as a way of tracking fluctuations in LFP dynamics. It consists of dynamic feature extraction and dynamic classification. Dynamic feature extraction determines the features and channels that indicate the best separability between classes of interest (disease and non-disease states) such that they can be adopted for use. And dynamic classification uses model selection to select the most compatible classifier for use. This research objective is achieved in Chapter 4.

- To help in overcoming the power and bandwidth constraints imposed by wirelessly monitoring biomarkers in PD patients, an on-site real-time hardware-efficient PD

detection processor is developed. It was prototyped on a field programmable gate array (FPGA) platform so that it can be made flexible enough to be updated as more knowledge regarding PD and DBS is obtained. It is ideally designed to monitor PD events to assist external caregivers or for use with stimulation devices. To obtain on-chip power and area estimates that could guide future developments, various implementations of the processor were mapped to 45 nm CMOS technology. This research objective is achieved in Chapter 5.

- As the need for a fully implantable aDBS system is becoming ever more important, a robust and resource efficient closed-loop control strategy adopting a *critic-actor* control approach was proposed and its feasibility investigated. The critic-actor configuration models a physician. The critic observes the symptoms against which disease progression are assessed and the actor takes possible actions to mitigate the symptoms – which are what a trained clinician does. This research objective is achieved in Chapter 6.

- Taking into consideration the shortcomings of implants for chronic DBS such as electrode migration, preliminary work on an output stage that could facilitate directional steering is presented. Directional steering has been suggested as a possible solution to overcome spatially dependent changes in chronic DBS. This research objective is achieved in Chapter 7.

Other research gaps targeted are:

- To provide a direction for multidisciplinary researchers and stakeholders in the DBS community, a review highlighting the major challenges in DBS both from the point of view of engineers and clinicians is conducted. This research objective is achieved in Chapter 2.

- To assist with providing balanced methods for comparing brain-machine-interface (BMI) technologies, which aDBS systems are part of, the need for standard as well as custom metrics for assessing their performance are highlighted. This research objective is achieved in Chapter 3.

The long term goal is to realise fully implantable aDBS systems: for both closed-loop DBS and spatial adaptation. For that to be achieved, efficacious circuits and systems with ultra-low power implementations are required. The above objectives are a precursor to fully implantable aDBS systems.

1.3 Publications

The work reported in this thesis has resulted in the following publications:

- **A. Mohammed**, M. Zamani, R. Bayford, and A. Demosthenous, "Toward On-Demand Deep Brain Stimulation Using Online Parkinson's Disease Prediction Driven by Dynamic Detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2441-2452, Dec. 2017.

- **A. Mohammed**, V. Valente, R. Bayford, and A. Demosthenous, "Output stage of a dynamic current steering deep brain stimulator," in *2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, 2015, pp. 81-84.

- **A. Mohammed**, M. Zamani, R. Bayford, and A. Demosthenous, "Patient specific Parkinson's disease detection for adaptive deep brain stimulation," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 1528-1531.

- **A. Mohammed**, R. Bayford, and A. Demosthenous, "Towards Adaptive Deep Brain Stimulation for Parkinson's Disease: A Review" Submitted to *Neurodegenerative Disease Management.*

Other potential publications that could result from this work are:

- **A. Mohammed**, R. Bayford, and A. Demosthenous, "A Framework for Adapting Deep Brain Stimulation Based on Parkinsonian State Estimates," Manuscript in preparation for submission to *IEEE Transactions on Neural Systems and Rehabilitation Engineering.*

- **A. Mohammed**, R. Bayford, and A. Demosthenous, "Complementary Detection for Hardware-Efficient On-site Monitoring of Parkinsonian Progress in Patients," Manuscript in preparation for submission to *IEEE Transactions on Biomedical Circuits and Systems.*

- **A. Mohammed**, R. Bayford, and A. Demosthenous, "Measures and Metrics for Evaluating Detection and Classification Algorithms in Brain Machine Interfaces," Manuscript in preparation for submission to *Journal of Neural Engineering.*

1.4 Thesis Overview

This thesis consists of eight chapters. Chapter 1 has given a brief introduction on clinical DBS and the limitations that motivated the work in this thesis. Also, it details how this work intends to address the highlighted limitations to contribute to existing research in aDBS.
Chapter 2 provides background and associated literature on aDBS. It details the fundamentals of PD, commercial DBS systems, and the future potentials of DBS systems. Then it describes the major areas in which advancements are needed to facilitate aDBS. It goes on to describe the major advancements in two forms of aDBS: closed-loop DBS and spatial adaptation. The major findings in the chapter are summarised at the end.
Chapter 3 delves deeper into the major challenges of physiological signal analysis. Since closed-loop DBS is an instance of BMIs, it extends the review to BMIs. The chapter focusses on the analysis of sensing techniques and feedback signals that can be used for closed-loop DBS with particular emphasis on neural signals. It goes further to analyse the principal metrics and measures that can be used to evaluate efficacy, complexity and efficiency of feedback algorithms in closed-loop DBS and BMIs. Afterwards, it

provides a perspective on the need for standardising and/or customising performance metrics for feedback algorithms in BMI, such that algorithms could be assessed in a more objective, fairer and more balanced way. The main points in the chapter are highlighted at the end.

Chapter 4 addresses the first objective of this thesis through an exhaustive analysis of various combinations of PD detection algorithms consisting of feature extraction, dimensionality reduction and classification stages. The analysis was concerned with hardware-aware implementations; as such, trade-offs in relation to accuracy and complexity were paramount. A new dimensionality reduction method, the MRM, is proposed as part of the algorithms to be evaluated. The chapter goes further to propose a method for semi-synthetic LFP generation to enable validation of algorithms. It also demonstrates the use of dynamic detection schemes as a way to overcome the inconsistencies in feedback signals across time and patients.

Chapter 5 addresses the second objective by providing a hardware-efficient platform for on-site processing of neural signals such that PD events can be detected and remotely monitored at a reduced power and bandwidth requirement. The system is the first of its kind for on-site PD monitoring and detection. The system was prototyped on an FPGA platform and mapped to 45 nm CMOS process for insights into its microchip implementation in terms of power and area consumption.

Chapter 6 addresses the third objective by proposing and exploring the possibility of implementing a critic-actor control approach as a robust and resource efficient closed-loop control strategy for adapting DBS. The system leverages on a machine learning model as the critic and a fuzzy controller as the actor. The approach was adopted in an attempt to individualise therapy since there is still limited understanding regarding the mechanisms of DBS as well as the progression of PD in patients.

Chapter 7 addresses the fourth objective by proposing and presenting a neuro-stimulation technique in which stimulation field spread beyond targeted areas can be mitigated using dynamic current steering. The chapter focused on the design of an output stage for dynamic current steering. The output stage uses dynamic current sources (DCS) in implementing 8 push-pull current sources that are interfaced to 16 electrodes so as to enable current steering. The dynamic current sources operate in two phases - programming and stimulation. This technique was adopted because it reduces power consumption and output current mismatch. The 16 electrodes can be driven both as anodes and cathodes and it is made to be scalable for use with more electrodes. The circuit was implemented in a 0.18 μm CMOS technology.

Chapter 8 summarises the main contributions of the work and provides future directions in which the algorithms and implementations in the work can be developed towards the goal of realising a fully implantable real-time aDBS system.

Chapter 2

Towards Adaptive Deep Brain Stimulation for Parkinson's Disease

Adaptive DBS presents a promising therapy for PD. However, implementing an effective adaptive system for DBS has not been realisable. This has partly been due to the poor understanding of the cellular mechanisms responsible for the network effects of DBS and PD. This chapter presents a comprehensive literature on background and other works related to aDBS. It starts with the fundamentals of DBS. Then delves into the various processing stages required to realise an aDBS systems; which is the section on "Towards Adaptive DBS". With a firm understanding of the fundamentals, the next two sections highlight the advances in closed-loop adaptation and spatial adaptation in DBS. The next section highlights the major developments in the field of DBS. And the final section summarises the major points in the chapter.

2.1 DBS Fundamentals

The basal ganglia are a group of subcortical nuclei targeted by clinical DBS. They are responsible for many functions including, posture and movement control. So far, many studies have confirmed the predominant claims that the basal ganglia exercises a major role in movement control. Their findings established that dopamine deficiency in the basal ganglia disrupts the discharge patterns in its nuclei like the globus pallidus externa (GPe), the STN, the GPi and the SN. This disruption results in an imbalance between inhibitory and excitatory outputs which leads to the onset of PD symptoms [40]. Generally, more insights into the physiological causes of PD are still gained. To alleviate PD symptoms, clinical DBS is used. Even though the mechanisms by which this happens are still under debate. Clinical DBS systems are mainly marketed by three companies; Medtronic, Boston Scientific and St Judes Medical. The major innovations by manufacturers of clinical DBS systems have centred on increasing the specifications for stimulation parameters; with work on aDBS systems still at the research stage. Nevertheless, DBS has established itself as the preferred therapeutic intervention during the advanced stages of PD. This is why its therapeutic benefits have been explored in more than 40 brain sites for about 30 clinical disorders [41]. Even with all these, approval has mainly been for its use in three stimulation sites: STN, GPi and Vim.

And for only three conditions, namely, PD, dystonia and tremor.

2.1.1 The Basal Ganglia

The basal ganglia is located in deep regions of the brain and comprises of a group of subcortical nuclei, which are directly responsible for posture and movement. The basal ganglia is notably involved in PD. PD primarily affects the SN (shown in Figure 2.1). This decreases the production of dopamine, an inhibitory neurotransmitter, which interferes with movement execution in patients. Apart from the SN, other nuclei of the striatum, the globus pallidus (GPi and GPe), and the STN. The largest nucleus is the striatum consisting of the caudate nucleus and the putamen, and it contributes to mainly motor and cognitive functions. The pathological changes in the basal ganglia play a major role on conditions like slowing of movement, abnormal body postures, involuntary movements, and/or a combination of these abnormal conditions [42]. The mechanisms of the basal ganglia that lead to PD are described in the next section.

2.1.2 Parkinson's Disease

The fundamental networks involved in normal and PD pathophysiology are compared in Figure 2.2 [3]. Generally, the region below the cortex, which is the basal ganglia, is mainly involved in movement regulation.
In Figure 2.2, the normal (left side) and PD (right side) networks of the brain are depicted. Excitation is represented in red and inhibition in blue. Thickened lines indicate increase in excitation (red) or inhibition (blue), while the broken lines represent reduction. The major points to note between PD (right) and normal (left) conditions are:

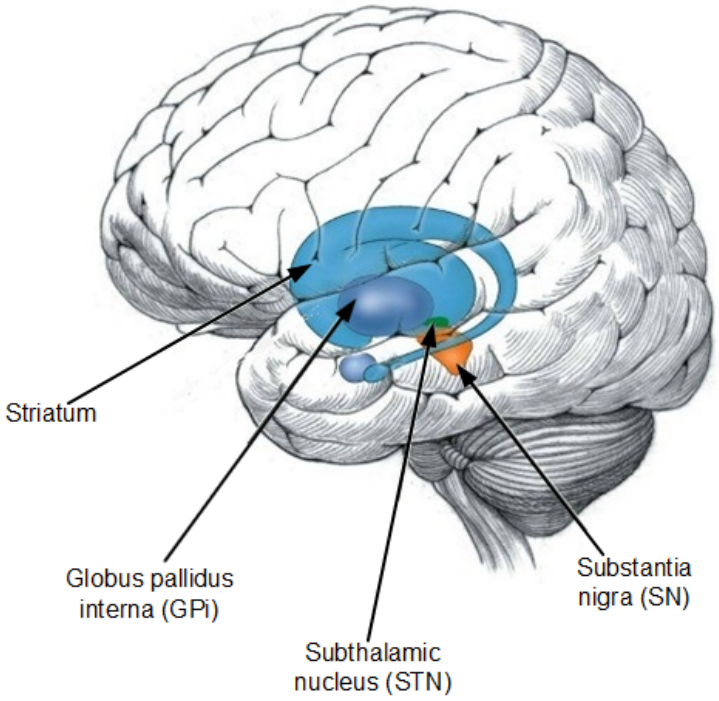


Figure 2.1: Internal structure of the basal ganglia (adapted from [2]).

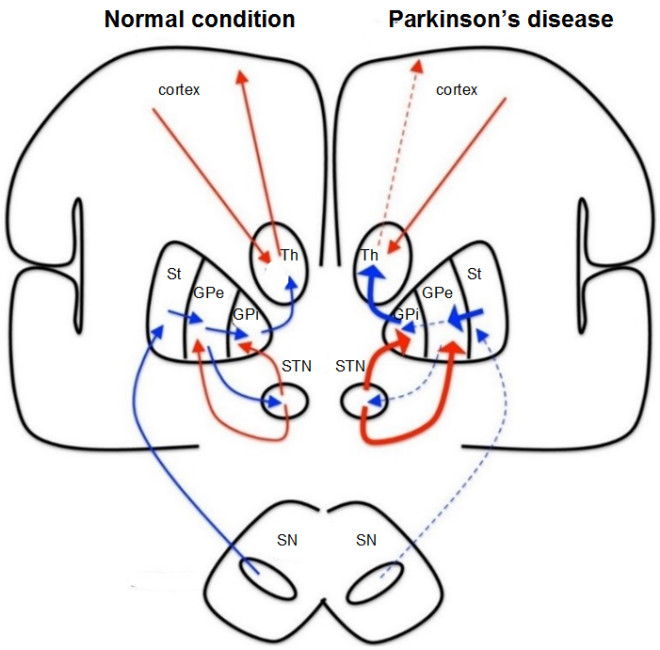


Figure 2.2: Schematic of network imbalance in PD, contrasting non-PD and PD brain. Excitation is shown in red and inhibition in blue. Thickened lines indicate increase in excitation (red) or inhibition (blue). St is the striatum, GPe is the globus pallidus externa, GPi is the globus pallidus interna, Th is the thalamus, SN is the substantia nigra and STN is the subthalamic nucleus (adapted from [3]).

- The SN is the principal area of concentration of the dopamine-containing neurons.
- The striatum receives reduced inhibitory inputs from the SN due to PD.
- This reduced inhibition from the striatum is passed on to the GPe and the GPe passes it onto the STN.
- The reduced inhibition from the GPe to the STN results in increased excitatory outputs to the GPi and GPe.
- The GPi then increases its inhibitory outputs to the thalamus, which has direct connection to the cortex and motor areas of the brain. This imbalance in PD accentuates the increase in inhibition from GPi to thalamus.
- This creates a phasic and oscillating inhibition from the GPi to the thalamus, which manifests as PD symptoms in patients. PD therapy is used to decrease this excessive inhibition moving out of the GPi into the thalamus.

Generally, PD results from a decrease in neural output of dopamine in the SN. As can be seen in Figure 2.2, the reduction in the concentration of dopamine disrupts the network balance between inhibition (blue lines) and excitation (red lines) in the network [3]. Disruptions within the interconnections of the nuclei in the basal ganglia can result in a number of movement disorders which can lead to extreme movement impairment and uncontrollable involuntary movements. Other major diseases instigated by this disruption include Huntington's disease, ET, TS, dystonia and many others. The

dominant symptoms in PD includes resting tremor, bradykinesia, postural instability and muscular rigidity. The first line of action on diagnosing PD is pharmacological therapy, mostly levodopa (L-dopa). L-dopa is a chemical building-block that is converted into dopamine by the body, thereby replacing some of the dopamine lost as a result of PD. Since PD is an incurable and progressive disease, therapy is only palliative. Initially, therapy results in gradual improvement in motor functions; however, motor impairments becomes resistant to L-dopa therapy, and this leads to additional symptoms like dyskinesia [43]. This stage is called advanced PD, and other clinical interventions like DBS may be required to mitigate PD symptoms.

2.1.3 DBS Systems

The main components of a DBS system are the intracranial electrode and implantable pulse generators (IPG), which are connected through an extension wire as shown in Figure 2.3.

The surgical procedure for DBS implantation of a DBS device involves two stages. The first stage is the DBS lead placement and the second is IPG placement. The Vercise neurostimator in Figure 2.3 uses the Boston Scientific lead model 2201 with lead dimensions as follows: diameter (d) is 1.3 mm, inter-electrode spacing (s) is 0.5 mm and electrode length (l) 1.5 mm. The Medtronic 3389 compatible with Activa neurostimulator has dimensions, d = 1.27 mm, s = 0.5 mm and l = 1.5 mm. While the St Jude 6149 model has dimensions, d = 1.41 mm, s = 0.5 mm and l = 1.5 mm. DBS lead placement begins by defining targets, common targets are STN, GPi and VIM. DBS leads are implanted through a burr hole ranging from the cortex to the basal ganglia as is shown in Figure 2.4.

After DBS lead implantation, optimal stimulation target is identified and patients are assessed for clinical benefits as well as side effects. When the DBS lead is firmly in the target position, the IPG is implanted below the clavicle and connected to the leads

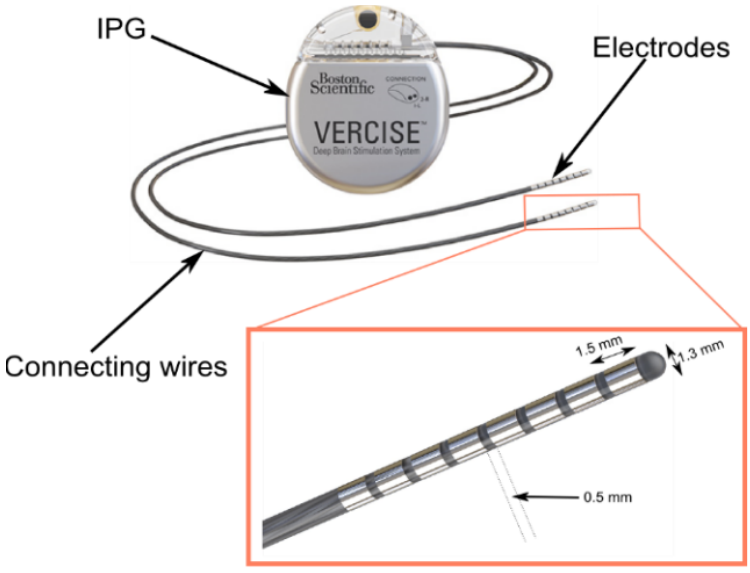


Figure 2.3: Boston Scientific's Vercise neurostimulator (adapted from [4]).

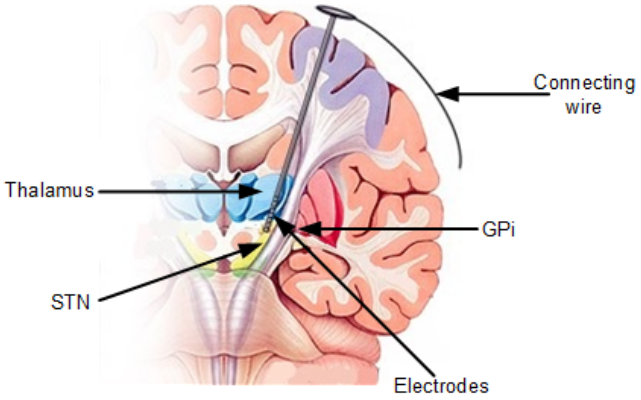


Figure 2.4: DBS lead placed in STN (adapted from [5]).

through the connecting wires [44].
The operating principle for devices produced by all manufacturers is the same; however, there are slight differences in technical features. The characteristics of the main DBS systems marketed by the major manufacturers are summarised in Table 2.1 [45–49].

2.2 Towards Adaptive DBS

There are currently two main methods DBS clinical outcomes can be adapted: closed-loop DBS and spatial adaptation in DBS. Closed-loop adaptation in DBS involves, automatic and personalised adjustments of stimulation parameters in response to brain signals. This is so that therapeutic benefits can be enhanced. Alternatively, therapeutic benefits can also be enhanced by adapting to changes in ageing-related lead migration, which will be termed spatial adaptation in this literature. So far, primary focus in DBS adaptation has been placed on closed-loop DBS.

This section will focus on the major areas that are necessary in driving development in aDBS, particularly for closed-loop DBS implementations. To enable advancements in closed-loop DBS, development in four major areas are required: understanding the underlying mechanisms of the brain, how useful information can be obtained from neural dynamics through analysis, how the information obtained can be used to trigger stimulation and how therapeutic stimulation affects brain tissue. This section is structured to touch on the four highlighted parts as shown. The typical processing chain of a closed-loop DBS system highlighting the major sections is shown in Figure 2.5.

2.2.1 Spatio-Temporal Neural Dynamics

The brain is populated with neurons whose temporal and spatial dynamics are very complex. Understanding the brain dynamics is key to understanding PD and other neurological disorders. Brain dynamics result from the interaction of millions to billions of neural cells. The neurons are the fundamental computational and signalling units in the brain. Figure 2.6 shows the structure of a neuron. A neuron consists of soma, axon and dendrites.

Table 2.1: Programming settings and characteristics for different DBS systems.

Manufacturer	IPG Model	Stimulation Parameters		
		Frequency (Hz)	Pulse width (μs)	Amplitude
Medtronic[a] [e]	Activa PC[b] [g]	30 -250	60-450	0 - 10.5 V 0 - 25.5 mA
		2-250		
	Activa RC [b] [c]	30 -250		0 - 10.5 V 0 - 25.5 mA
		2-250		
	Activa SC[d] [g]	30 -250		0 - 10.5 V 0 - 25.5 mA
		2-250		
Boston Scientific [a]	Vercise [d] [f]	2 - 255	10 - 450	0 - 20 mA
	Vercise PC [f] [g]	2 - 255	20 - 450	
	Vercise Gevia [d] [f]	2 - 255	20 - 450	
St Jude Medical [e]	Libra [e] [f]	2 - 240	50 - 500	0 - 12.75 mA
	Libra XP [c] [e]	2 - 240	52 - 507	
	Brio [c] [d]	2 - 240	52 - 507	
	Infinity [e] [g]	2 - 240	20 - 500	

[a] Uses both current and voltage sources.
[b] Bilaterally implanted leads with 4 electrode contacts/lead.
[c] Rechargeable battery.
[d] Unilaterally implanted leads with 4 electrode contacts.
[e] Has multipolar configuration. In addition to its polarity (tripolar or quadripolar), multipolar configuration can be used in uni/bipolar polarity. Medtronic stimulators in current source configuration are limited to only uni/bipolar polarity.
[f] Bilaterally implanted leads with 8 electrode contacts/lead.
[g] Non-rechargeable battery.

The soma contains the cell nucleus. The dendrites consists of the branches that enable connections between neighbouring cells. The axon is the cable-like structure that transmits nerve signals to target cells via synapses. The synapses allows the passage of electrical or chemical signals to neighbouring neurons. The axon is made up of a myelin sheath, which consists of dielectric materials that aid the propagation of impulses. The

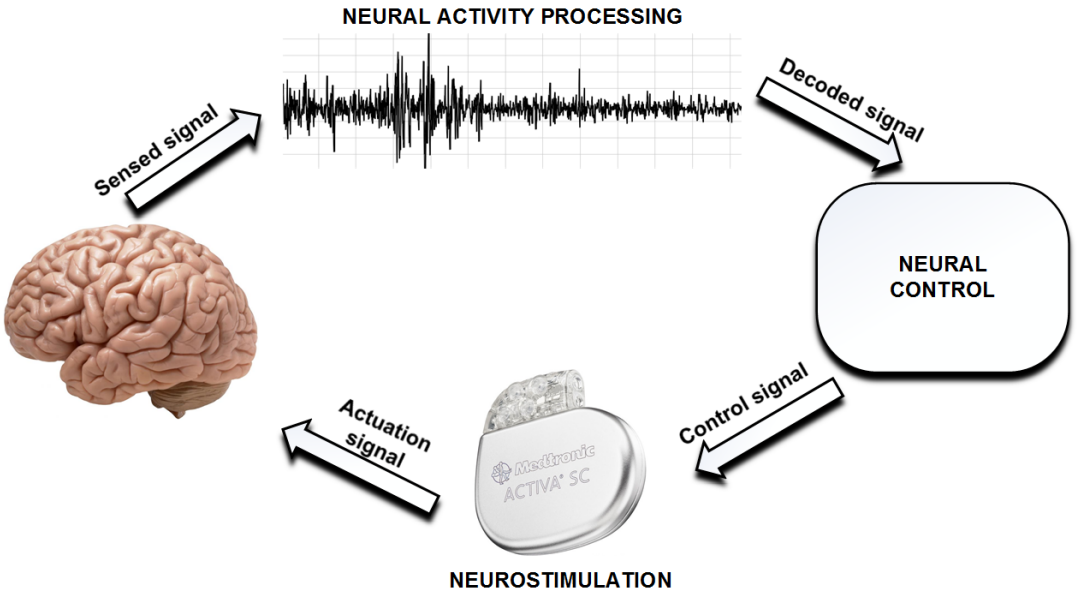


Figure 2.5: Typical processing chain of a closed-loop DBS system.

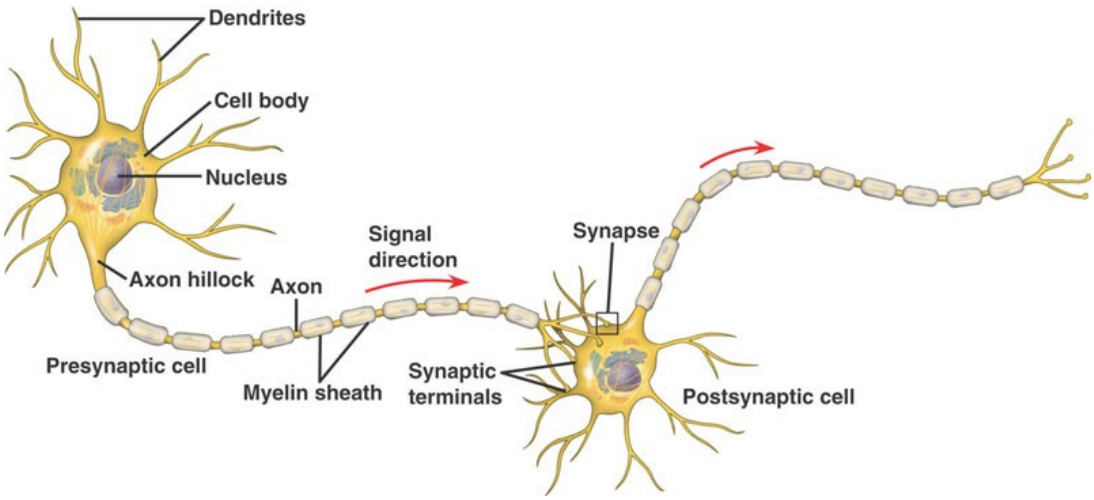


Figure 2.6: Structure of a neuron.

axon has segments that are divided by regions known as nodes of Ranvier, which amplify and transmit signals down the axon. Neurons are organised into layers in the brain, and they number around 10^{11} in the brain with each one having about 10^4 connections. As such, modelling the complete dynamics of the brain is a near-impossible task. However, various attempts at modelling the brain dynamics have been made, notably are the single cell models by Hodgkin and Huxley [50]; to the more detailed single cell models by FitzHugh and Nagumo [51]; and to the aggregate level models by Wilson and Cowan [52]. All hypothesise that the brain has non-linear dynamics. And like all non-linear systems, the individual dynamics of its neurons varies from the superposition of the dynamics of the individual neurons. The Wilson and Cowan model is the baseline for describing spatio-temporal neural dynamics and several extensions of it have been used in modelling neural dynamics [6].

The Wilson and Cowan model suggests that all neural processes are dependent on the interaction between inhibitory and excitatory cells. Wilson and Cowan in their work [52], found out that interaction between excitation and inhibition in neural cells was necessary in order to establish dynamic stability in the brain. This led to the premise that excitatory-inhibitory interplay was crucial to restoring network dynamics in epilepsy [53]. This principle of balancing in neural dynamics by inhibition and excitation was further affirmed by experimental studies [54, 55]. The Wilson-Cowan model asserts that there are separate populations of neurons that are excitatory and inhibitory. The models were made general enough to encompass all possible scenarios of inhibitory and excitatory interaction. This was done by setting connection strengths and parameters of the model that represented all possible cortex topologies. Mathematically, the Wilson-Cowan equations for spatially homogeneous cases can be summarised as follows [56],

$$\tau \frac{dE}{dt} = -E(t) + (1 - rE(t)) f_E[w_{EE}E - w_{EI}I + h_E(t)] \ , \tag{2.1}$$

$$\tau \frac{dI}{dt} = -I(t) + (1 - rI(t))f_I[w_{IE}E - w_{II}I + h_I(t)] \ , \tag{2.2}$$

where $E(t)$ is the proportion of excitatory cells active per unit time, $I(t)$ is the proportion of inhibitory cells active per unit time, $h_E(t)$ is the external input to the excitatory population, $h_I(t)$ is the external input to the inhibitory population and f_E and f_I are the firing rate functions (are sigmoid functions) for the excitatory and inhibitory population. Also, W_{EI} and W_{IE} are the connection between excitatory and inhibitory populations, with W_{EI} representing excitatory to inhibitory, while inhibitory to excitatory is represented by W_{IE}. In addition, there is interaction within the excitatory and inhibitory subpopulation and these are represented by W_{EE} for self-excitation and W_{II} and self-inhibition. And finally, r represents the absolute refractory period. The model assumed neurons to have three different connections: excitatory, inhibitory and self-feedback connections. A schematic of this is shown is shown in Figure 2.7. Figure 2.7 (a) maps aggregate populations of neurons on a grid, with each cell on the grid (aggregate population of neurons) having independent excitatory (green coloured) and inhibitory (red coloured) neural populations with activity levels $x_i(t)$ and $y_i(t)$ respectively. The activity levels quantify the proportion of firing neurons in each region at time t. Also, Figure 2.7 (b) shows the different configuration of connections, connection strengths and parameters. The left region, a, receives excitatory (green arrows), inhibitory (red arrows) and self-feedback (blue arrows). It shows weak (dashed arrows) and very weak arrows (dotted arrow) originating from the right region b. The alphabets above each arrow indicates synaptic weights. The inhibitory and excitatory network connections involved in PD pathophysiology explained in Section 2.1.2, were inspired by neural connections as illustrated using the Wilson and Cowan model [52].

The Wilson and Cowan model has gained wide acclaim because it introduced population thinking and statistical theory to analyse brain activity. Its analysis captures

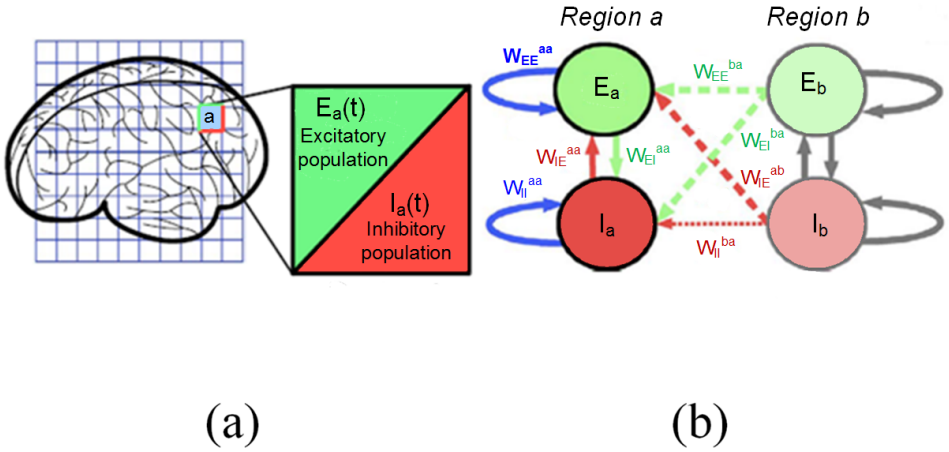


Figure 2.7: Schematic illustration of neuronal dynamics. (a) Showing the cortical region divided into a grid of macroscopic regions. Each region i (blue squares), comprises of excitatory (green) and inhibitory (red) neural population. (b) Showing all possible configurations of neural connections (adapted from [6]).

the mean field activity by showing non-linear wave propagation and non-linear pattern formation. Consequently, extensions of it have been used to explain the effects of anaesthetics, seizures and various stimuli to the brain [3]. The model uses numerical solutions and phase plane methods to describe the responses of neuronal population to stimuli. The model established the existence of multiple stable states, oscillations, spatial patterns, traveling waves and hysteresis in neuronal response. Aside from spatio-temporal dynamics, brain activity measured from several techniques (measuring techniques are discussed in Chapter 3), are shown to contain frequency specific oscillations [57]. These macroscopic activities (involving between $10^8 - 10^{11}$ neurons) are representative of high-level functions like movement, cognition, memory or perception. In order to understand the mechanisms involved in such tasks, analysing and interpreting spectral information is necessary. This is explained further in the next section, which provides some background on neural signal analysis.

2.2.2 Neural Signal Analysis

In monitoring changes in sensory, motor and cognitive tasks, neural signal as well as external body signals are potentially useful [58]. A major challenge of closed-loop DBS is the choice of a suitable feedback signal. Below are some of the requirements necessary for feedback signals in PD [16]:

- The signals should be bio-markers that reliably reflect all symptoms of the disease, impairment or disability (such as tremor, bradykinesia and rigidity across patients).
- Relationship between signals and impairment should not be correlative, but should be causative.
- The measure should also be instantaneous such that therapy does not lag impairment.
- Invasiveness of the recording technique should be reduced to the barest minimum.

Based on these, neural signals are the most advantageous for use as feedback signals. They can be sensed using an implanted custom-integrated chip that allows for measurements, processing and analysis [59]. The same leads can be used both for stimulation and measurements. In addition, they are desirable because signals that are obtained closer to the stimulation sites are more reflective of patient condition [9]; making signals obtained from deep regions of the brain more desirable. The type of feedback signal used determines the bio-signal processing technique adopted. Nonetheless, in most neurodegenerative disorders like PD and epilepsy, biomarkers have a fairly similar signal processing chain. Figure 2.8 shows the typical signal processing chain for neural signal processing consisting of feature extraction, dimensionality reduction and classification. The first step in neural signal processing is the transformation of acquired physiological data normally in time domain, to a computationally efficient form for further processing. This transformation could be from time domain to a more efficient time domain

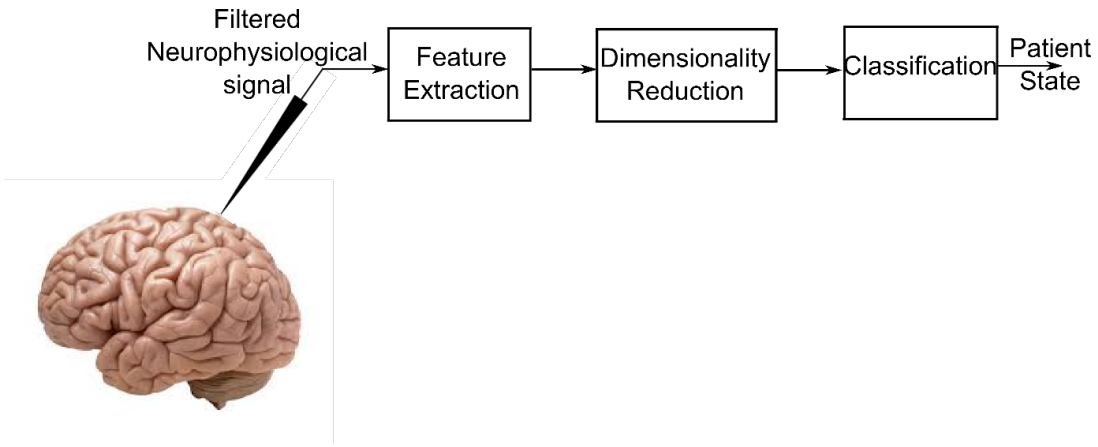


Figure 2.8: Typical neural signal analysis chain used in closed-loop neural systems.

form, from time domain to frequency domain or from time domain to time-frequency domain. Time domain to time domain transformation may be necessary because relevant information may not be obvious from individual samples, however more detailed information can be obtained by filtering, smoothing or averaging recordings to create epochs that make the onset of various events more obvious. In contrast, frequency domain features may be required since neuronal networks often use frequency coding to communicate [60–62]. Hence transformation of time series data into frequency information may be necessary in applications were neuronal firing rate signify the occurrence of certain events. Nevertheless, time or frequency information alone may not be sufficient in real-time detection of pathological states [9]. Thus, techniques that provide both time and frequency information may be required. After transforming recorded data into a useful form, an additional processing stage targeted at reducing redundancy, increase processing speed and providing computational efficiency is normally used. This additional processing is termed dimensionality reduction (DR). DR involves reducing the number of features that will be used for the detection or classification of patient states. In patient state detection, high dimensionality in feature space increases data overfitting, which results in poor generalisation and inefficient detection of patient state. In brain-machine-interface (BMI) applications, the most primitive dimensionality reduction technique is manual feature selection. Manual feature selection is implemented by selectively retaining the most relevant features. Kung and colleagues [63], proposed feature selection in genomic signal processing where by redundant genes were identified so as to sparsify the feature vectors. In another study on epilepsy seizure detection using electroencephalography (EEG) signals, Lee et al. [64], adopted feature selection by channel reduction. Channel reduction was implemented by manually identifying redundant channels, after which they were eliminated. In closed-loop DBS, neuronal information is used in regulating stimulation. This involves extracting disease relevant features and classifying the patient state using the extracted features in order to adapt stimulation. Generally, high-order classification models are required due to the complex nature of the underlying processes in physiological signals. These high-order models increase the complexity of the computations involved; which in turn makes clas-

sification incur the most energy across the bio-signal processing chain. The complex nature of the underlying processes in physiological signals have led to the development of various machine learning frameworks for classification. These frameworks are not only concerned with detecting patient states, but can also be used in understanding the evolution of the patho-physiological processes in patients; thus modelling transitions between various states in a disease.

2.2.2.1 Challenges of Neural Signal Analysis

The development of fully implanted PD detection processors is difficult due to the computationally intensive nature of neural signal processing. This forms a major bottleneck in developing closed-loop implantable DBS systems for clinical intervention. The need for fully implantable PD detection processors is pushing research into the investigation of resource efficient algorithms and techniques. For hardware platforms targeting conditions with highly unpredictable physiological signals like PD, high-order models are required. However, for fully implantable systems, the power density for tissue damage limits the ability to use complex models which can be computationally intensive. Some of the common challenges when dealing with physiological analysis in BMI applications are highlighted below.

- Correlations of physiological signals to clinically relevant states are hard to model. High-order data-driven models may be required to distinguish disease states of interest from non-disease states. This makes current *feedback algorithms* inefficient in tracking patient states.
- Identifying precise correlations are difficult because the physiological manifestation of disease and non-disease states vary from patient to patient. This imposes a wide range of specifications on *sensing devices*.
- *Feedback signals* are inconsistent across time, this makes static feature extraction unreliable.

The ability to assess physiological signals over a large number of channels will be essential to model their correlations with disease states. It is for this reason that data-driven approaches like machine learning are emerging as powerful tools that could be used to handle this challenge [65,66]. The complexity of brain signals on its own makes closed-loop DBS very difficult to achieve. The following sections expatiate on some of the highlighted challenges, as encountered by current state of the art neural signal processing systems.

Feedback Algorithms: One of the major issues in achieving closed-loop DBS has been insufficient data relating to various patient states. This has led to poor mapping between disease state and biomarkers. These poor mapping coupled with the dynamic nature of PD biomarkers, have resulted in suboptimal feedback algorithms that inefficiently track patient states. Thus, having very clear insights into the underlying mechanism of PD at the neurological level is necessary to facilitate the design

of feedback algorithms tailored to suppress PD. In other to facilitate the mapping of stimulation parameters and biomarkers, several algorithms have been used. The work in [67] uses search algorithms, adaptive controllers were used in [58, 68] and statistical methods were used in [69] used to identify the relationship between stimulation and resulting LFP output. These are computationally intensive methods that are not practical for fully implantable closed-loop DBS systems due to high power and bandwidth requirements. Also, they are mainly simulation studies that use assumptions that may not be practical for real life implementations. Essentially, poor insights into the internal workings of DBS has inhibited the development of effective feedback algorithms that can be used in tracking disease states.

Sensing Devices: One major challenge is the type of recording device used and the noisy nature of recordings. Noise sources originate mostly from the electrode-tissue interface (ETI) and can degrade signal processing and analysis capabilities [70]. Devices susceptible to noise have a tendency to obtain recordings that obfuscate disease states even in conditions with more distinguishable neuronal activity like epilepsy. This can be more challenging in disorders with less distinguishable disease states. For specific patterns to be obtained through neural signal analysis, sensing devices that are less susceptible to surrounding noise are required. In disorders whose mechanisms are still under debate, the large variability in the physiological manifestation of disease and non-disease states across patients imposes wide specifications on sensing due to the need for an increased bandwidth, dynamic range and so on; making the needs of various patient difficult to fit on a single device.

Feedback Signals: Primarily, feedback could be implemented using neurophysiological signals or external body signals. Neurophysiological signals have proven to be the most suitable due to their high reliability [16]. Neurophysiological signals can be invasive and non-invasive; non-invasive measurements are obtained at the stimulation site. Due to the dynamic nature of PD, recordable signals showing direct correlation to patient states are yet to be established. Even though correlation has been established in some patients, enough evidence identifying certain signals as the cause of PD have not been established. Thus, correlations could be as a result of secondary effects of the disease. The dynamic and unpredictable nature of candidate signals for PD detection makes establishing a suitable feedback signal a major challenge. Subsequently, this has led some studies to suggest the possibility of combining more than one neurophysiological signals [16]. Others have proposed the use of multiple features from the same neurophysiological signal [58, 71, 72]. While many have suggested complementing internal body signals with external body signals. This is in order to maintain consistency in feedback signals. However, adopting some of this proposals can make analysing physiological signals computationally intensive. As a result, robust signal analysis are needed to adapt to the changes in the feedback signals. Based on this, having a single feedback signal as the universal biomarker may not be a sustainable approach towards closing the loop.

2.2.2.2 Prospects of Neural Signal Analysis

Due to the complex nature of physiological signals, current laboratory-based closed-loop DBS systems use multiple external computers to process sensed signals. However, clinically viable realisations of closed-loop DBS needs to be more robust and autonomous such that all processing is implemented on-site and real-time – this will require high functionality on-chip processors. The shift towards on-chip processing has necessitated the need for simple but efficient processing techniques in order to reduce computational complexity so that processors can be realised in CMOS technology within the power consumption constraints. The bulk of the power consumed in bio-signal processors for closed-loop DBS is incurred at the classification stage. Less computationally intensive classifiers can be adopted if more accurate biomarkers for PD are identified. Current bio-signal processing approaches use supervised machine learning methods which require the use of labelled data for training. In the future, bio-signal analysis for closed-loop DBS has the potential of utilising unsupervised machine learning techniques for classification, which will create more robust algorithms that can handle the complex and dynamic nature of the electro-physiological signals. Online unsupervised machine learning techniques have been pioneered in spike sorting and other BMI applications [73]. The total shift towards unsupervised learning techniques will eliminate the need for time-stamped measurements, thus totally eliminating the intervention of trained operators (or clinicians). The first step towards achieving this is by adopting adaptive algorithms which respond autonomously to changes in physiological signals, as is part of the objectives of this work. The prospects of bio-signal processing in closed-loop DBS depends on power consumption constraints and the available insights into the DBS and brain mechanisms. Better control over these will facilitate the implementation of effective, autonomous and efficient processing techniques that can be used to adapt stimulation. To overcome the high efficiency and efficacy demands required in order to close the DBS loop, effective control strategies can complement the processors to satisfy the imposed resources constraints. It is for this reason that the next section will focus on control strategies for neural systems.

2.2.3 Neural Systems Control Strategies

Neural systems control is a highly multidisciplinary field that uses feedback signals from nervous systems to control functionality. Because of how they are implemented, they are widely called closed-loop or feedback control. It has been widely established that closed-loop systems achieve better efficacy and efficiency than their open-loop counterparts. This has been one of the driving forces in the promotion of closed-loop neuro-prosthetic systems. Closed-loop neuro-prosthetic systems are envisioned to complement, restore and improve neural functions like sensory, motor and cognitive tasks. At present, many of the systems are proof-of-principle, and transition to clinically approved interventions are still hampered by issues like: establishing relationship between acquired signal and patient condition, invasiveness of recording devices and how chronic implantation can be sustained in invasive devices. To circumvent this, most studies have opted to use models with limited detail to describe the basal ganglia, central nervous system

(CNS) and so on. This is understandable because of the many inconsistent theories describing the mechanisms of PD and DBS [3, 30–34]; developing a reliable model becomes difficult. With no accurate patient models, current closed loop strategies use a trained clinician to close the loop by adjusting stimulation parameters based on visual feedback from patients. However, parameters obtained under this condition may not guarantee sustained therapeutic effects [74]. More so, the procedure is costly and time consuming. In addition, only a very small proportion of stimulation parameter space can be explored using trained clinicians in closing the loop. The tediousness of this process makes it essential to implement control strategies that facilitate parameters adjustment in real-time based on some quantifiable (and possibly objective) measure. This reduces the risk and tediousness of constant clinical intervention. The difficulty in producing accurate models that will provide insights into closed-loop control and the difficulty associated with frequent programming of stimulation parameters makes the implementation of closed-loop control a non-trivial task. Nevertheless, several attempts at closed-loop control have been made. The first step in closing the loop is to identify, acquire and analyse biomarkers so that they can be used as controller inputs, which was covered in the previous section. This is necessary because the efficacy of a controller is partly dependent on the quality of its control input. The following sections describe the most prominent control techniques that have been adopted in neural systems.

2.2.3.1 Open-loop Approach

This is the only non-feedback control approach. It generates control signals based on input to the plant without recourse to the output from the plant. An open-loop control strategy assumes that the system has enough knowledge to provide corresponding output with specific input signals. This assumptions is faulty and misleading and cannot hold in complex systems. Nevertheless an open-loop system is always the first step in controller design. All of the currently marketed DBS devices for PD use open-loop control.

2.2.3.2 Bang-band (Simple Feedback) Approach

A bang-bang controller commonly known as an on-off controller (or hysteresis controller), is a feedback controller that switches abruptly between two levels. Bang-bang control is a closed-loop control technique that uses threshold crossing to directly generate the control signal to the controller without prior pre-processing. These are normally used in low complexity systems. In closed-loop DBS applications, they have mostly been used as the main control strategy in a number of experimental studies [24, 61, 74–82]. Nonetheless, they can been used as complementary controllers to assist more complex controllers [58]. With more discrete levels, they can be used with more classical controllers like the Proportion-Integral-Derivative (PID) controller. In this configuration, it uses error signals as input to the controller. The proportional, integral and derivative inputs are calculated from the systems response to its inputs. In PD, using a single feedback signal (for example beta band activity) may be suitable for controlling bradykinesia and rigidity, however this may not be effective for tremor [25]. This approach may be sub-optimal considering how PD consists of various underlying

symptoms including bradykinesia, rigidity and tremor. The bang-bang approach uses one-dimensional feedback signals. Many studies in closed-loop DBS have focussed only on the beta-band activity as input to the closed-loop controller. Using more discrete levels as well as multi-dimensional feedback signals presents a feedback approach more sophisticated than the bang-bang, which are described in the following sections.

2.2.3.3 Internal Model-Driven Approach

This is a feedback control approach that incorporates a model of the system, typically a black box model which is defined based on the input-output relationship of the system. This uses system identification to establish input-output relationships for a system that may be difficult to model. At every instant, before providing control commands, the model provides a prediction of the system behaviour and controller inputs are determined based on a cost function that determines optimum parameters. Examples of these include recursive autoregressive models and Kalman filter methods. For DBS, some computational models have implemented closed loop control using these methods [69, 83].

2.2.3.4 Classifier-Driven Approach

This uses a feedback control approach that relies on mapping between discrete states to determine input signal to controller [84]. Classifiers can use supervised, partly supervised or unsupervised machine learning methods to obtained input-output relationship from existing. These relationships are used to generalise the relationship on unknown data [85]. The unsupervised systems segregate the system into states with similar patterns using only input data without labels. Typical examples are techniques like k-Means [10] , and the O-sort algorithm used for clustering in spike sorting. The partly supervised learning algorithms have training data that is partly labelled. For supervised learning, the main methods are k-nearest neighbours (k-NN), support vector machine (SVM), logistic regression (LR), linear discriminant analysis (LDA), artificial neural networks (ANN) and a host of others. When used with a controller, it is typically used like a bang-bang controller (but with prior pre-processing and multidimensional feedback signals) since it has binary classes representing healthy and unhealthy states. Unlike the bang-bang controller, it uses multi-dimensional feedback signals to cater for all symptoms [64]. It has found wide application in epileptic seizure detection [64, 86]. For epilepsy, an FDA approved closed-loop system, Responsive Neurostimulator System (RNS), marketed by NeuroPace Inc. uses this approach for modulating therapy for drug-resistant epilepsy [86].

2.2.3.5 Actor-Critic Approach

The actor-critic approach models the relationship between the physician and the automated neuromodulation system. The critic (like the "trained clinician") assesses the state of the system based on a cost function and provides the information to the actor. The critic learns about the system by studying its input and resulting output responses. The actor, unlike the error signal of other control techniques provides control signal based on evaluation from the 'informed critic'. This method has been adopted for

neural control in [58], and is gaining wide-spread acceptance for controlling non-linear systems because it adopts a technique that resembles real-life clinical interventions. Table 2.2 summarises the advantages and disadvantages of the various neural system control strategies.

Table 2.2: Advantages and Disadvantages of Neural Systems Control Strategies.

Control Strategy	Advantages	Disadvantages	DBS related study
Open-loop	It is simple, straightforward and easy to implement.	It is unrealistic and misleading. It cannot correct for errors and cannot compensate for disturbances.	Conventional DBS.
Bang-bang (Simple Feedback)	It is simple, straightforward and easy to implement.	It can be unstable as it monitors fluctuations between two pre-set levels.	[24, 61, 74–82].
Classifier-Driven	Provides a simple approach for cases with an established relationship between discrete disease states and therapy.	Classification algorithms can be computationally intensive and directly mapping states to therapy may not provide therapeutic benefit due to the non-linearity of the disease/disorder.	[87–91].
Internal Model Driven	Input-output model may assist in adapting to changes in dynamics of disease or disorder.	They are yet to be attempted in any investigative studies. They have only been used in computational studies.	[69, 83].
Actor-Critic	It adopts techniques that incorporate real-life clinical diagnostics and intervention.	Requires very accurate sensor that captures and tracks real-time biomarker fluctuations, so that they can be "critiqued" to enable immediate action. State estimates may not be representative of symptom severity.	[58]

2.2.4 Neural Tissue Stimulation

DBS pacemakers consists of a small battery-powered IPG (also known as pacemaker) implanted under the skin near the clavicle with leads that attach to electrodes placed within the brain at specific locations depending on the condition the device is used to treat; PD (electrodes located at STN or GPi) or ET (electrodes situated at VIM). The IPG continuously delivers low intensity, high frequency electrical pulses to parts of the brain. The following section gives a brief overview on the physiological processes resulting from therapeutic neural stimulation and the available stimulation techniques.

2.2.4.1 Neural Tissue Excitation

Brain tissue consists of electrically excitable neurons. The first electrical model of a neural cell was developed by Hodgkin and Huxley in 1952 [50]. This modelled the physiology of the axon membrane using an electric circuit. The model is shown in Figure 2.9. In the model, the ionic current I_{ION} flowing through the membrane is split into

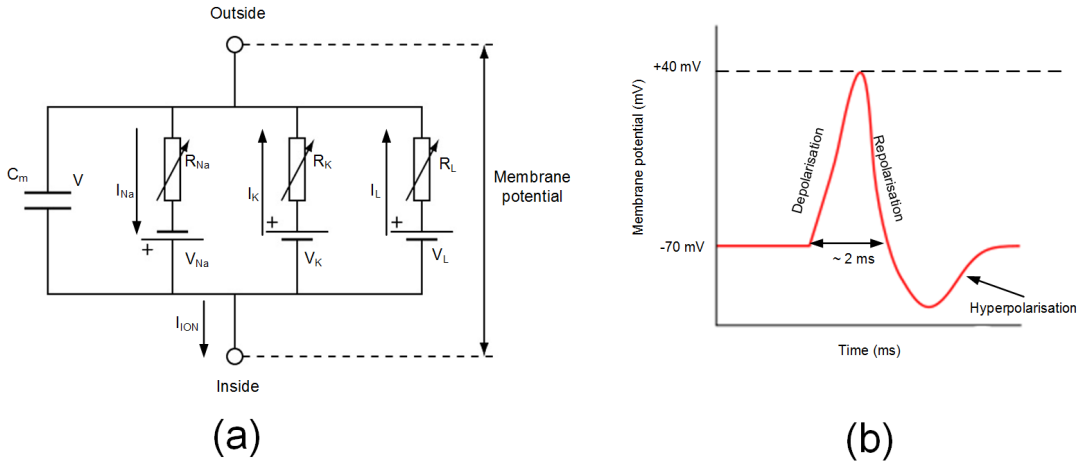


Figure 2.9: Hodgkin and Huxley model and action potential. (a) Electrical model of a neural cell (b) Propagated action potentials; theoretical (top) and experimental (bottom).

three components; potassium current (I_K), sodium current (I_{Na}) and leakage current (I_L). Each cell is represented by a membrane capacitance (C_M), a resting potential (V), three ionic channel potentials (V_K, V_{Na} and V_L), and channel resistance (R_K, R_{Na} and R_L). This model describes the characteristics of propagated action potentials; their shape, amplitude, refractory period, excitation threshold and ionic channels. Primarily, electrical stimulation generates action potentials by delivering current into a cell such that its membrane voltage exceeds the threshold voltage, which results in an action potential. This is termed cell membrane depolarisation and is depicted in Figure 2.10. The rate of change of the generated membrane potential based on Hodgkin and Huxley model is [92],

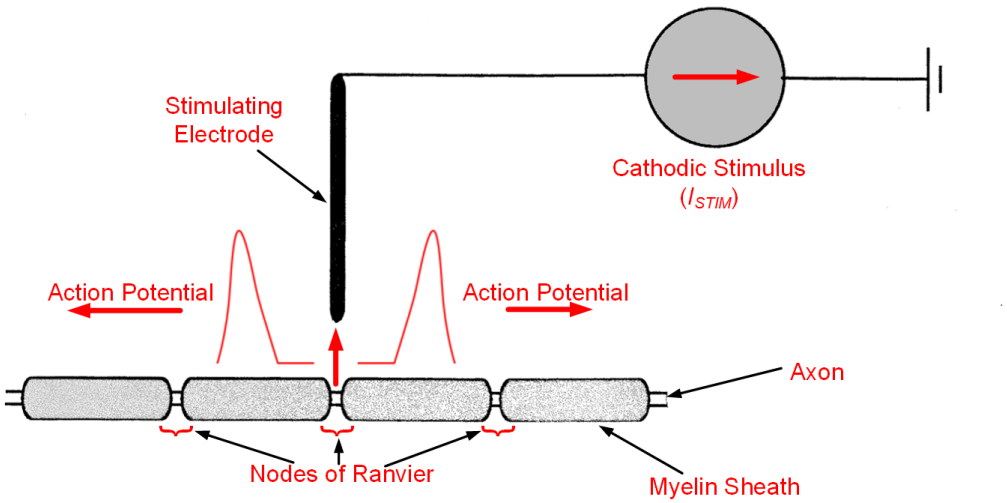


Figure 2.10: Cell membrane depolarisation by stimulation (adapted from [7]).

$$\frac{dV_m}{dt} = -\frac{-I_{ION} + I_{STIM}}{C_m} \tag{2.3}$$

I_{ION} is is the ionic current in the neuron resulting from sodium, potassium and leakage currents in the neuron. V_m is the membrane voltage, I_{STIM} is the applied stimulation current to the cell. This injects a charge during stimulation, considering I_{STIM} is a square pulse. The total charge injected is,

$$Q_{tot} = I_{STIM} \cdot T \tag{2.4}$$

where T is the pulse width for a mono-polar stimulation. To ensure that stimulation is safe, a biphasic stimulation protocol needs to be adopted so that zero charge is accumulated during stimulation to avoid irreversible phenomena like tissue damage and electrode corrosion, that result from charge accumulation [93]. Biphasic stimulus paradigms have become the adopted standard. They provide stimulation using a sequence of two cycles of different polarity, so that the charge injected during the first cycle is removed during the second cycle. Additionally, the maximum safe electrode charge density is not exceeded [1].

2.2.4.2 Therapeutic Stimulation Techniques

Therapeutic stimulation has been in existence for several centuries. It has been used to mitigate sensory deficits such as blindness, deafness, chronic pain, urinary incontinence, paralysis, PD, ET among others [95]. In therapeutic brain stimulation, the resulting effects depend on the stimulation site, stimulation parameters and uncontrolled biological effects [22]. In therapeutic DBS, it is still not clear which neural elements are targeted. Nonetheless, basic knowledge regarding the responses evoked by various stimulation modalities is needed in order to adopt a technique that optimises therapy, minimises side effects and maximises battery life. The common stimulation techniques are voltage controlled stimulation (VCS), current controlled stimulation (CCS) and charge controlled stimulation (ChgCS). Efficiency and safety are the key factors that result in the selection of one of these stimulation techniques. VCS has been found to be a very efficient method due to its long battery life span. However, its safety has been questionable, due to the fact that constant voltage excitation is delivered to an ETI with a variable impedance. This variation will lead to the accumulation of charge at the ETI. Generally, the charge at the stimulation site is desired to be within a safe limit known as the water window [93]. If the charge exceeds this limit, irreversible tissue damage or electrode corrosion may occur. Conversely, CCS offers better control over injected charge than VCS. Even though it incurs extra power consumption due to the conditioning circuit used in producing constant current excitation. This impedes efficiency. As a trade-off between efficiency and safety, ChgCS has been introduced using switched capacitor based stimulators [96]. Table 2.3 summarises the various stimulation techniques and their characteristics.

[1] Conventional DBS electrodes made of platinum-iridium have an electrode charge density of $30\mu C/cm^2$/ phase [94]

Table 2.3: Benefits and Limitations of Therapeutic Stimulation Techniques.

Stimulation Techniques	Benefits	Limitations	Example Commercial Devices
VCS	Provides the longest battery life span compared to other techniques.	Possible accumulation of charge at ETI raises serious safety concerns.	Mainly used by Medtronic devices: Activa PC, Activa RC and Activa SC.
CCS	Offers better control over injected charge than VCS.	Extra power consumption is incurred in the conditioning circuit used in producing constant current excitation.	Commercial devices are experiencing a shift towards CCS. Most commercial devices in market have CCS mode operation.
ChgCS	Offers a trade-off between safety and efficiency.	The need for charging capacitors could make them bulkier than necessary.	Still at infancy. More research is still required before it is commercially used.

Clinical DBS design is tending towards current controlled stimulation due to its safety. This has resulted in a surge in the design and development of current neuro-stimulators, although their therapeutic benefits over VCS are yet to be clinically proven [97]. The choice of the stimulation technique impacts on the lifespan of the battery. With the advent of rechargeable batteries, the emphasis on power consumption has reduced [25]. Nevertheless, some patients are unsuited to rechargeable IPG systems [98]. Even patients who can tolerate the use of rechargeable batteries, would want to recharge as infrequently as possible. The charging process is tedious and patients may be required to recharge the system every week [99]. If the battery is allowed to completely run down three times, then replacements may be required, which means a new surgery. The need for low power stimulation techniques is ever more important considering the demand for cranially mounted IPGs where patients undergo only a single procedure, compared with the current two stage procedure for electrode and IPG placement [100,101]. Therapeutic stimulation techniques are selected based on a trade-off between efficacy, safety and power consumption.

2.3 Advances in Closed-loop DBS

DBS provides an overall improvement in Parkinsonian patients' quality of life. Over time, the efficacy of the therapy is reduced due to disease progression, environmental factors, mechanical factors, and behaviourally induced changes in network activity [35]. This is mainly as result of the ability of nervous systems to adjust their activities in response to new situations or to changes in their environment, which is termed neuro-plasticity. Additional programming sessions may be required to manually adjust stimulation settings [35]. Current DBS systems are poorly suited to cope with the dynamic nature of PD. This has led to growing interest in the design of closed-loop systems for adapting DBS. In order to implement closed loop systems, existing open

loop systems can be optimised by incorporating feedback schemes. However, the current dilemma has been the poor correlations between feedback signals and the motor score measured through the Unied Parkinson's Disease Rating Scale (UPDRS), even though some correlations between motor states and certain feedback signals have been identified. These inconsistencies necessitates the need for adaptive feedback schemes through closed-loop DBS. Closed-loop DBS involves adjusting stimulation parameters to characteristic changes in biomarkers. Biomarkers for closed-loop DBS are mainly neurophysiological signals and external body signals. Neurophysiological signals consists of single-unit recordings, multi-unit recordings, local field potentials (LFP), global field potentials and Neurochemical recordings. On the other hand, external signals are mainly from electromyographt (EMG) and accelerometer signals. Recently, various studies in closed-loop DBS have focused more on computational models. However, very little work has led to clinical studies. Below is a brief overview of different attempts at closed-loop DBS for PD patients.

2.3.1 Experimental studies

Clinical studies in closed-loop DBS have been carried out using Parkinsonism induced animal models and in patients with informed consent. These studies have used neurophysiological and external signal as biomarkers. Using spike activity, the work in [75] investigated closed-loop DBS using 1-methyl-4-phenyl-1,2,3,6tetrahydropyridine (MPTP) induced primate model of PD. The results showed that stimulation triggered by spikes from primary motor cortex (M1) and GPi, led to a reduction in GPi firing rates, which were representative of improved condition. The stimulation was triggered based on a predefined delay of 80 ms. The study presented results that bettered those of continuous stimulation. This provided an interesting proof of principle, however it is not clear how the technique will fare across patients and over time. Also, Brittain et al. [76], proposed another approach of delayed-stimulation, however, alternating current transcranial stimulation was used instead. An extension of this technique using external body signals was presented by Cagnan et al. [77]. It uses a unique stimulation approach that selectively controls neural synchrony by delivering stimulation to the ventrolateral thalamus. This was timed according to the patient's tremor rhythm. The study claimed that DBS can be precisely timed to disrupt disease pathophysiology. It worked on the premise that stimulation selectively regulates neural synchrony through phase-specific stimulation [80]. It attained promising results in six of the nine subjects used. DBS delivered with personalized frequencies has been reported to effectively reduce UPDRS motor scores [102]. A different study [24], uses the beta band local field potentials (LFP) in triggering stimulation on threshold crossing in eight patients. They reported 56% reduction in stimulation time and reduction in the energy requirement of the closed-loop DBS compared to its open-loop counterpart. In a follow-up study to overcome the shortcoming of the previous study, which were short sessions (10 minutes) and unilateral stimulation. Little et al. [78], used 4 PD patients with bilaterally implanted DBS for a sessions lasting 2 hours, there was a marked improvement in motor score as against the open loop approach like in [24]. The major limitations of

both studies were; the thresholds were set heuristically and provisions were not made to adapt them to changes in patient conditions as feedback signals are reported to be inconsistent across time and patient [16]. This static thresholds adopted made the system unadaptable to drift in neuronal characteristics. More so, in another study, LFP's were also used as biomarkers for detecting seizure like activity in [58]. Seizures were induced by high amplitude current stimulation in ovine models, and seizure like activity was detected using LFP power from pre-defined frequency bands. These power measurements were classified with a Fisher discriminant and an actor-critic control policy was used to regulate stimulation. This approach does not make adequate provision for the dynamic nature of the LFP signals. In an almost similar fashion, Johnson and colleagues illustrated the promising utility of closed-loop DBS for PD based on STN beta LFP levels [79]. Stimulation was delivered only when STN LFP beta activity was elevated. The study used a female rhesus macaque monkey induced by MPTP. However, it concluded that closed-loop DBS systems may need alternate and complementary biomarkers and/or algorithms to reach their full therapeutic potential [79]. Similarly, LFP's were used in [87] for tremor detection in PD patients using a multi-layer neural network for classification. This method presents a good concept that could be used for demand driven stimulation. However, its use of static detection scheme coupled with additional complexity introduced by multi-layer neural networks, which could make it difficult for full implantation, since the neural-networks are trained offline. The drift in characteristics of neuronal signals over time and across patients, necessitates the use of a tracking paradigm that will adapt to this changes in biomarkers. Using microelectrode recordings (MER), Kostoglou and colleagues [88], proposed a random forest approach for identifying UPDRS improvement in PD patients undergoing DBS – off dopaminergic medications. The findings suggested that electrophysiological signals had a strong correlation to improvements in UPDRS score, and they could be used to predict UPDRS score. Major breakthrough in DBS adaptation can be achieved if a relationship between motor improvement and electrophysiological signals is achieved. Alternatively, in terms of using external body signals, the work in [89] used a combination of surface EMG (sEMG) and accelerometer signals to predict tremors in PD patients. Even though the study predicts tremor with substantial accuracy, patient distress in PD ideally precedes symptoms manifestation; which makes the technique unsuitable for adaptive DBS therapy. Generally, feedback using external body signals has an accompanying delay in detecting PD coupled with the discomfort associated with externally attached sensors. This makes them not very viable for use as biomarkers. In contrast to previous works, attempts using neurochemical biomarkers have been made. Grahn et al. [74], describes a rat model that uses fast scan cyclic voltammetry (FSCV) to determine evoked dopamine in the striatum, after which stimulation is regulated accordingly. Another study that uses neurochemical responses from rat models is the Mayo Investigational Neuromodulation Control System (MINCS) [103], it is designed to interface with FSCV. The system recorded striatal dopamine release in anesthetised rats and stimulation was wirelessly regulated in response to evoked neurochemical sig-

nals. Studies using neurochemical changes offer improved selectivity, sensitivity and specificity; however, they are not suitable for real time applications due to the poor temporal resolution associated with neurochemical sensing. Also, the need to miniaturise implantable devices makes its use unsuitable, as neurochemical sensors tend to be bulky. Though they tend to have a better consistency across patients compared to other biomarkers. Table 2.4 highlights some of the characteristics of the various experimental studies.

Table 2.4: A summary of experimental studies using closed-loop DBS.

Ref	[75]	[76]	[77]	[24]	[78]	[58]
Experimental subjects	2 African Green Monkeys	12 Patients with tremor dominant PD.	9 patients (6 had ET and 3 had dystonic tremor (DT))	8 PD patients	8 PD patients	An ovine model
Stimulation site and paradigm	Pulse train at 130 Hz delivered to GPi and motor cortex (M1) after 80 ms of detecting single unit neural activity.	Transcranial alternating current stimulation (TACS) delivered to M1 on detecting cortical activity.	Accelerometer triggered stimulation delivered to Ventrolateral thalamus based on the patient's tremor rhythm (3-8 Hz).	Stimulation delivered to STN ramped on threshold crossing of beta-band LFP for each patient.	Bilateral stimulation of STN triggered in response to beta-band LFP threshold crossing.	Network activity in broad beta band (10 - 30 Hz) in Hippocampus (HC) and anterior nucleus (AN) used to trigger stimulation based on an actor-critic control policy.
Outcome	There is more than 50% alleviation in occurrence of tremor in subjects.	Achieves almost 50% average reduction in resting tremor amplitude.	Between 8% - 51% tremor suppression in dominant tremor axis for ET patients.	56% reduction in average stimulation time	The mean improvement in UPDRS scores for unblinded is 66.2% and 49.7% for blinded assessment.	Seizures suppressed within a second. And uses about 10% of nominal power over a period of 15 months.

Ref	[79]	[87]	[89]	[88]	[74]	[103]
Experimental subjects	A female rhesus macaque (Monkey)	7 PD patients	8 patients with tremor (4 ET and 4 PD)	20 PD patients	Four anesthetized rats	Three anesthetized rats
Stimulation site and paradigm	Stimulation triggered in STN at threshold crossing of LPF beta band.	LFP recorded from STN (No stimulation).	sEMG signals used (No stimulation).	LFP recorded from STN (No stimulation).	A look-up table was used to determine the stimulation parameters based on evoked dopamine (neuro-transmitter) release. Sensing through the Striatum (St) and stimulation at the medial forebrain bundle (MFB).	Stimulation wirelessly ramped in response to neurochemical recording. Sensing in St and stimulation at MFB.
Outcome	Stimulator is switched on about 50% of the time but performance comparable to continuous stimulation.	55.2% - 100% tremor detection accuracy.	The predictor achieves an average sensitivity of 100%, with a mean accuracy of 85.7% for all ET trials and 80.2% for all PD trials.	The UPDRS improvement (good or poor) for 19 of the 20 patients were correctly classified.	There is a reported increase in neurotransmitter, which is suggestive of improved motor conditions.	As in [74], there is an increase in striatal dopamine release.

2.3.2 In-silico Studies

Even though the underlying principle of PD remains unclear, various computational studies have been based on neuronal models. Notable among these are the phase dynamics model, firing rate models and the stochastic models. The earlier models at closed-loop DBS have used phase evolution of neuronal cells to model PD dynamics. In the phase dynamics model, STN neurons fire in an uncorrelated (desynchronized) way under healthy condition, and PD onset results in increased synchronised activity of the neurons. Using this theory, Rosenblum and Pikovsky [80, 104]; proposed a technique whereby time delayed stimulation is used to suppress self-synchronisation in a neuronal population. The assumption is that DBS pulses (with a fixed offset) are used to disrupt the oscillatory activity as a result of PD onset. In another study using the phase dynamic model [81], spatially distributed stimulations were used in addition to time delayed stimulations. This was to alter neuronal synchronisation by superimposing out of phase signals so as to desynchronise the local neural population. Lysyansky and colleagues also presented a study using phase dynamics model [61]. Here, multi-site stimulation for desynchronising local neuronal population was used, by stimulating at different sites using the resonant frequency of a local neuronal ensemble. This was in contrast to the conventional multi-site approach. Additionally, using the phase dynamic model, closed loop desynchronising strategy in [105] uses a DBS signal in proportion to the mean field of the neuronal population. Drawing upon the theory that desynchronising stimulation protocols are effective as closed-loop techniques for the control of synchronisation in ensemble of neurons, Popovych and colleagues used pulsating closed-loop DBS to restore the desynchronising characteristics in the STN-GPe neuron model [106, 107]. The amplitude of the stimulation pulses was defined with respect to the phase delay in the LFP feedback signals. Generally, studies using the phase synchronisation approach are theoretical in nature and have had some experimental validation [75–77]; which have been found to be effective within individual patients. A major shortcoming of implementing the phase dynamics model will involve an understanding of the optimal way to keep pathological neurons effectively desynchronised. In addition, the experimental validation of the phase dynamics have used stimulation for periods not longer than 30 seconds, this puts to question its performance across time and across subjects. Most recent studies have used stochastic models and firing rate models. Using a stochastic model, neuronal activity is modelled using various stochastic processes, ranging from regular, irregular, random and bursting neuronal activities. Santaniello and colleagues [69], using system identification, modelled a system that automatically adjusted stimulation based on LFP recordings from the stimulation electrodes. It models PD as a stochastic process having regular, irregular, random and bursting STN neurons. It then uses a recursive autoregressive model to identify the relationship between stimulation parameters and the stochastically modelled LFP recordings. In a simulated PD state, stimulation parameters were determined using the initially established relationship between LFP power spectrum and stimulation. In a similar vein, Liu et al. [83], uses system identification to establish the relationship

between stimulation and neuronal responses. Unlike in [69], which was more concerned with generating the stimulation pattern that restored neural activity to PD-free state. The work by Liu and colleagues chose the stimulation pattern that restored neural activity with the least energy expenditure. However, both models in [69,83], model the behaviour of the basal ganglia network using less than a hundred neurons. This makes them inadequate for producing the complete dynamic behaviour of PD. Nevertheless, they give an insight into the possibility of using feedback from LFP and other neural activities to restore patient condition if there is an established relationship between changes in LFP, patient state and stimulation parameters. Other studies have used firing rate models to characterise the excitation state of different basal ganglia neuronal populations. This was modelled in [82] using the average rectified value (ARV) of beta band LFP oscillations in firing neurons. Firing rate models are ideally suited for control theory, as they use the state excitation of different basal ganglia regions. In addition, Brocker et al. [108], used genetic algorithms to optimise the temporal pattern of stimulation in a computational model using a firing rate model. This resulted in a low frequency stimulation for the model and this was validated on a rodent (using 45 Hz stimulation). Likewise, Brocker et al. had proved in an earlier study the temporal efficacy of properly timed DBS in PD [109]. In [110], firing rate models were used to explore the use of classical feedback control algorithms for adapting stimulation. The best controller for suppressing neuronal synchrony was obtained for a configuration using a full PID control law. Table 2.5 summarises the merits and demerits of computational studies in closed-loop DBS using different neuronal models.

Table 2.5: The major neuronal models used in computational models for closed-loop DBS.

Type of Model	Merits	Demerits	Ref.
Neuronal phase models	Disrupts self-synchronisation in neuronal population using time-delayed stimulation.	Requires multi-site stimulation in order to disrupt synchronisation. In addition, there is no understanding of the optimal way to keep pathological neurons effectively desynchronised. There is still limited experimental studies that verify its efficacy.	[61, 80, 81, 104–106].
Stochastic models	Models neuronal dynamics using a range of stochastic processes ranging from regular, irregular, random and bursting neuronal activities; that are present within a neural population.	Mostly uses a small neural population. This may be inadequate and unrepresentative of the entire neuronal population.	[69, 83].
Firing rate model	Single neuron models of STN, GPe, GPe or thalamo-cortical neurons are connected to represent the basal ganglia network. This makes it more manageable and easy to undertake thorough control analysis.	It is still not clear if conclusions drawn from this models remain true if the number of neurons, their parameters or network connection are changed [111].	[82, 108–110].

Generally, simulation studies assess the possibility of implementing real life closed-loop DBS. Nevertheless, they incorporate many model assumptions whose validity might vary across patients. This makes their use limited, even though they assist in providing insight into how algorithms for adapting DBS can be implemented.

2.3.3 Peripheral Devices and IC Implementations

Current implementations of closed-loop DBS use software programs for signal analysis [24, 58]. They have high energy demands that could make them impractical and not portable for use. Incorporating low-power hardware or IC recording and processing can reduce some of the limitations associated with practicability and portability [112], which is necessary for fully implantable closed-loop systems. IC implementations have mainly focused on neural recording and stimulation stages that can be incorporated into a closed-loop DBS system. There are however peripheral devices developed as part of or towards incorporation into a closed-loop system. For ASIC development, [113] presents a closed-loop DBS system having 64 stimulating channels and 8 recording channels, which was tested in a rat model. The stimulators were implanted in STN of the rat and the recording electrodes were implanted in the motor cortex. The novelty in the system was the adoption of a logarithmic ADC (log-ADC) for data acquisition of electrophysiological signals. The log-ADC is used to implement high dynamic range neural recording covering spikes at 50 μV to LFPs at up to 5 mV. These recorded LFP and spike activities are used to trigger stimulation. The device consisted of off-chip processing of acquired neural signal via a wireless link. An extension of the work incorporates on chip

processing [114]. This is done using a logarithmic domain digital signal processor and a PI controller in order to control stimulation. Making the closed-loop system operate completely in log-domain; which simplifies multiplication and division operations into addition and subtraction respectively. The prototype systems present an implantable system-on-chip (SoC) demonstrating closed-loop DBS algorithms. However, the use of multiple locations for stimulation and recording makes the system invasive; an ideal closed-loop concept should work without the need for extra implants. Also, the system uses static features for detecting PD, which has no provision for the dynamic nature of PD. Conversely, concurrent sensing and stimulation was implemented in [115] with the prototype system implementing adaptive stimulation in response to induce seizure in an ovine model, nevertheless using static features. Salam et al. [116], implemented seizure detection and suppression on an FPGA platform. Seizures were induced in rats using intraperitoneal kainic acid injection. Only the population of rats responsive to the kainic acid injection were used for the study. The study found out that event-triggered stimulation reduced seizures by 90% compared to continuous stimulation which only reduced seizure occurrence by 17%. The results presented show the benefits of event-triggered stimulation compared to continuous stimulation. A major question the study fails to address is the possibility that seizure suppression could be as a result of some epiphenomena associated with stimulation spacing (as continuous stimulation has a tendency to induce seizures [58]) rather than event-triggered stimulation. Aside from the comparison with open-loop stimulation, comparison between intermittent stimulation and event-triggered stimulation would have more clearly demonstrated the superiority of event-triggered stimulation compared to other methods. For the most part, very little effort has been put in the design of application specific IC (ASIC) for closed-loop DBS. This has largely been due to the insufficient empirical evidence on the behaviour of the DBS mechanism.

2.3.4 Limitations and Future Directions

Experimental studies in closed-loop DBS are normally carried out within two to seven days after electrode implantation [25]. After electrode implantation, there is a reduction in PD symptoms, which is mostly as a result of stun effect. Stun effect results in a temporary reduction in parkinsonian symptoms and an unresponsiveness in LFP signals after electrode implantation [117]. This means improvements in patients condition may not be representative of the new therapeutic regimes, but as a result of stun effect. It is one of the reasons post-operative programming is delayed [118]. Experimental studies may not easily bypass the stun effect since its exact duration is still yet to be established [25]. Additionally, closed-loop DBS may need to be tested in a chronic setting, because the efficacy of conventional DBS has been reported to fall with time, which is primarily as a result of the habituation effect [21, 119]. Longer trials are necessary to determine if this is also the case for closed-loop DBS. Most of the experimental studies have used a bang-bang control approach and have mainly focused on monitoring beta band LFPs only [24, 82, 120]. However, using only beta band activity and setting heuristically obtained (static) thresholds may be ineffective and

suboptimal, as there are serious questions regarding their ability to track fluctuations across time and patients [24, 121]. They have been reported to correlate with symptoms in bradykinesia and rigidity; however, this is not the case for tremor. In tremor dominant PD, gamma [16, 122, 123], and tremor [124] band activities have been found to correlate with PD symptoms. These further questions the use of one-dimensional feedback signals in a bang-bang feedback approach. Multidimensional feedback signals using a simple but sophisticated control approach may be more viable for closing the loop. The main limitation of sophisticated feedback algorithms is their computational power needs. This could be offset by the less frequent stimulation required as a result of closing the loop. So far, closed-loop DBS has been hindered by the poor understanding of the underlying mechanisms that result in improved patient conditions. Better understanding of the underlying mechanism will create a more accurate mapping between disease states and stimulation parameters; making improved computational models attainable. It is important to have realistic computational models, which will significantly improve the performance of sophisticated closed-loop systems incorporating multiple functionalities needed for closed-loop DBS. These will in turn lead to the development of thorough clinical studies aimed at investigating techniques that optimise clinical benefits that are tailored to patient's needs. Tailoring therapy to patients needs could be advanced by adopting a similar technique as in cardiac defibrillators, which uses a combination of dual sensor technology. This technology combines a short-term sensor and a long-term sensor, in which the short term sensors track immediate changes from a selected biomarker, and the long term sensor tracks biomarkers with a slow response rate. Nonetheless, combining sensors with different rate responses requires adequate blending of respective sensor activities. Irrespective of the development in closed-loop DBS, significant advancement can only be achieved if there are strong multi-disciplinary collaborations between clinicians, engineers, statisticians, health care professionals, computer scientists, regulatory experts and most importantly, end-users.

2.4 Advances in Spatial Adaptation

Spatial adaptation allows for the variation of stimulation focus without a corresponding change in lead position. It is a form of aDBS that adapts to changes or inaccuracies related to lead position and/or poor stimulation focus in relation to targeted neural structures [99]. Poor stimulation focus has often affected the therapeutic benefits of DBS [26]. Directional steering of stimulation is the main form of spatial adaptation, and it offers a new dimension to DBS therapy by directional control of stimulation, in addition to the control of normal stimulation parameters like amplitude, pulse width and frequency. Precise neural targeting has been reported to improve the therapeutic window of DBS by reducing the threshold for beneficial effect and increasing the threshold for side effects [125]. This reduces the need for high amplitude stimulation. Figure 2.11 depicts how current steering can be used in providing stimulation focus in inaccurately placed or ageing-related lead migration [8]. While the deviations appear to be small; they can result in drastic reductions in therapeutic benefits accruable [126].

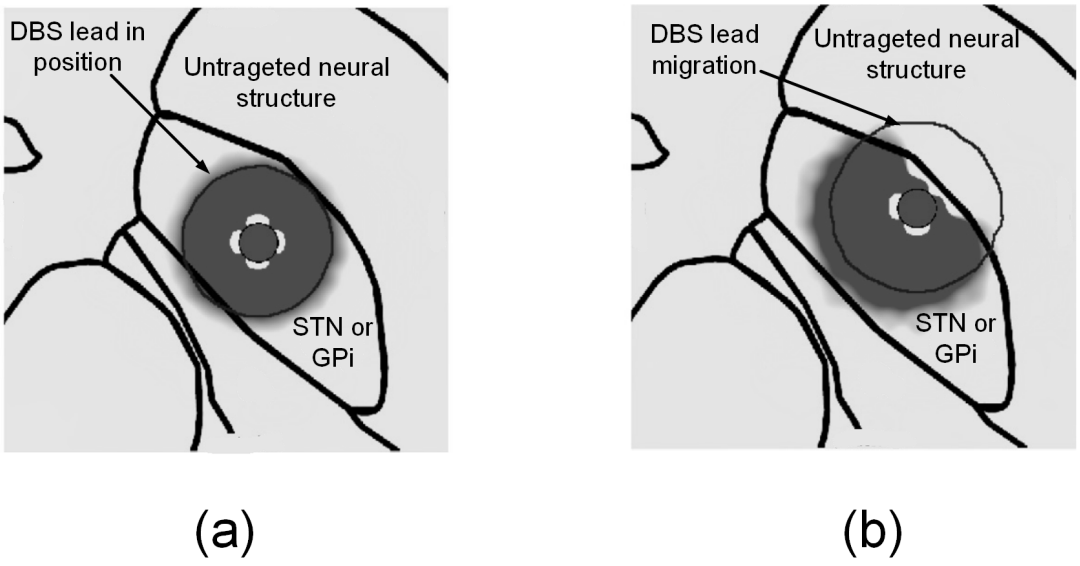


Figure 2.11: Computational model illustrating the effect of current steering on the volume of tissue activated (VTA). a) Showing accurately placed lead for STN stimulation. b) Depicts the use of current steering to control the activation field in inaccurately placed DBS leads for STN stimulation (adapted from [8]).

It is for this reason that the prospects of current steering have led to the commercial development of current steering systems [2]. The following sections will provide further information on directional steering in DBS and its prospects.

2.4.1 Directional Steering

Studies in directional steering have mainly focussed on the computational modelling of DBS leads and activation field spread, experimental studies, and CMOS implementation of current steering stimulators. Conventional DBS uses cylindrical electrodes which provide poor neural selectivity, since the stimulation distributes symmetrically around the electrode, thus targeting both intended and unintended areas. In order to overcome this, novel lead designs with electrodes having high contact resolution were modelled in [8]. The novel lead design uses 64 segmented electrodes which offer a larger electrode combination necessary for accurate field shaping and directional steering. The novel lead design shows a finer resolution in the generated potential distribution. This study motivated another study on current steering using 32-contact electrode leads [125]. The study uses eight patients to investigate the effects of using segmented electrodes on the therapeutic window of DBS, that is, the thresholds for beneficial effect in relation to the threshold for side effects. It was found out that the thresholds for side effects increased by more than 1 mA, and the thresholds for beneficial effect dropped by 0.5 mA, thus creating an increase in therapeutic window of about 1.5 mA. In terms of output stages for multipolar current steering in DBS, the first realisation was implemented in [127]. It uses a voltage controlled resistor (VCR) stimulation circuit and tri-polar current

[2] Boston Scientific, Sapiens SBS, and Aleva Neuroprosthetics have been aggressively involved in the development of current steering systems for DBS. Sapiens SBS was acquired by Medtronic, Inc. on Aug. 26, 2014 for approximately $200 million.

steering. The adopted tri-polar configuration has a tendency to be scaled to use higher number of electrodes. In another work, a CMOS circuit for current steering using multipolar and multisite current steering was also presented [128, 129]. So far, very little effort has been devoted to the development of output stages for current steering which are necessary to enable power management.

2.4.2 Benefits of Directional Steering

The cylindrical electrodes for conventional DBS were designed when there was very little or no scientific understanding of the mechanisms of DBS and neurodegenerative (and neuropsychiatric) disorders. However, with more insights gained, there is a need for explicit definition of spatial targets. This has resulted in a surge in the development of electrodes and leads for DBS [8, 130, 131]. These electrodes design have mainly focussed on two strategies for spatial adaptation [101]: 1) employing a number of small segmented electrodes which could be independently activated in response to issues like electrode migration in chronic DBS, or 2) customising a number of the cylindrical electrodes to improve stimulation of brain targets, especially in the more problematic regions. A typical example is in the stimulation of the Vim of the thalamus for ET patients. It has been established that stimulating the ventral caudal (Vc) nucleus of the thalamus induces parathesia [130]. Due to the proximity of the Vc to the Vim, there is the possibility that electric field spread could induce paraesthesia in patients [130]. This is why the major DBS devices manufactures are aggressively exploring the practicality of leads with provision for spatial adaptation [3]. Using segmented electrode provides more flexibility in administering stimulation, but increases the cost and complexity of the device. In addition, it imposes lower charge injection limits for safety and creates an impractical parameter search space for clinical DBS programming. These are the major issues in deciding which of the two possible routes (either using segmented or cylindrical electrodes) to follow in terms of spatial adaptation. Nevertheless, both techniques have shown that the benefits of accurate targeting and precise field control outweigh its shortcomings [101, 131]. Beyond the realm of stimulation, spatial adaptation was used to obtain a more precise biomarker sensing using high density electrodes [132]. Intraoperative electrophysiological measurements like LFP beta power, neuronal kinetic responses and neuronal firing have been used to identify suitable brain targets for neuro-motor disabilities [133, 134]. The sensed signal was used to trigger stimulation in a non-human primate. The results showed that high density electrodes had the potential to shape stimulation and biomarker sensing, which improved stimulation efficacy. Development of direction steering lead is necessary so as to optimise stimulation and biomarker sensing. The optimised stimulation/sensing offered by customised directional steering will go a long way in conserving energy, therefore increasing pacemaker battery life. As it has been highlighted, there are two main methods by which directional steering can be optimised: novel lead designs and energy

[3] Boston scientific developed "Vercise", a 16-electrode array current steering DBS system. Sapiens marketed the "SureSTIM" which has a 32-electrode lead (Sapiens has now been acquired by Medtronic, Inc). Aleva Neurotherapeutics developed the "DirectSTIM" which has 8-electrodes.

efficient output stages. Preliminary work on the latter is presented in Chapter 7 in an effort to realise a current steering output stage. Correspondingly, the manoeuvrability introduced by directional steering can lead to the exploration of potentially beneficial stimulation sites that were previously inaccessible by surgical techniques. Considering the effects of neuroplasticity, electrode migration and other extraneous effects, spatial adaptation of DBS has the potential to offer maximum therapeutic benefits without sacrificing efficiency.

2.5 Development Achieved in DBS

Research activity in the field of DBS has stagnated over the last few years [112]. Table 2.6 summarises the evolution and history of DBS and suggests possible future advances [1, 75, 99, 100, 135–137]. The major innovations in current DBS technology have mainly resulted in broadening the operating ranges of stimulation parameters; it is still not clear how this will increase clinical benefits. Nevertheless, it gives clinicians an increased flexibility and more degrees of freedom (DOF) to search for patient specific DBS parameters. On the other hand, increased flexibility also increases the economic cost of stimulation parameter programming due to the time-consuming trial-and-error process. Other innovations have been the introduction of rechargeable batteries and the concept of stimulation field shaping. Having rechargeable batteries is essential because the economic costs associated with DBS have been reported to be largely dominated by battery replacement cost [138]. As is normally the case in other therapeutic fields, not every innovation brings about considerable changes that may influence patients' quality of life, but many can improve safety, efficiency and flexibility both for patients and clinicians. Aside from incorporating additional functionality, technological advancements could manifest in the form of increased computing capability per chip at a reduced cost. Due to the short market cycle of electronic devices, continuous innovations are required to maintain market relevance even if they may have little or no benefit towards improving patients' quality of life. DBS is becoming more appealing due to the growing evidence pointing to the benefits of DBS at the early stages of PD [37]. Another contributing factor is the growing number of PD patients expected, which is expected to reach 8.7 million people worldwide by 2030 [139]. These could drive unit costs down as manufacturers achieve economies of scale. Low cost devices can translate to better market penetration particularly in developing countries. Currently, the annual sales estimates of DBS devices for PD is approximately $200 million to $300 million worldwide, but the coming years promise a further surge in sales [140]. It is estimated that more than 100 000 patients suffering from PD, pathological tremor and dystonia have been treated with DBS all over the world [141]. With the increase in the number of patients requiring DBS therapy, there is a need for smarter DBS programming strategies that can be self-optimising and autonomous. The thinking is that adaptive or smart DBS has great potential to keep DBS simpler (both for the patient and caregiver) and more viable. In epilepsy, a closed-loop Vagus Nerve Stimulator, the RNS NeuroPace, has been approved by the FDA for the treatment of refractory epilepsy [142]. Closed loop

therapy in epilepsy is easier than that of PD because non-healthy neuronal activity can be easily distinguished from healthy neuronal activity by trained clinicians, which is not the case in PD. Nonetheless, considering that both are closely related neurodegenerative disorders, it may not be long before fully implantable closed-loop systems are trialled. As things currently stand, researchers, entrepreneurs and other stakeholders in the DBS field believe that aDBS will be the silver bullet that will solve the myriad of problems currently associated with conventional DBS for PD.

Table 2.6: History and Evolution of DBS.

Period: Evolution of DBS	Subperiod (Quarters) Q1	Q2	Q3	Q4
1870 - 1969	Direct electric stimulation attempted on animals and then humans. Stereotactic frames to enable access to the brain developed.	–	–	Stereotactic frames are finally used for human surgery. Brain stimulation used to treat chronic pains and neuropsychiatric disorders. Intermittent chronic DBS trialled to reduce tremor in PD.
1970 - 1989	Stimulation used to treat chronic pain, epilepsy and other movement disorders.	Medtronic trademarks the term 'DBS' for neurostimulation. Stimulation of Periventricular-Periaqueductal Gray (PVG-PAG) is used to treat chronic pain.	Thalamic stimulation is used for treating tremor and dyskinesia.	UPDRS developed as a clinical assesment tool. Thalamic stimulation used for treating depression. Stimulation of the VIM is used for treating PD and tremor.
1990 - 2004	The efficacy of STN lesion in treatiing primates induced with movement disordes through MPTP. Implantable battery-driven DBS pacemakers.	STN stimulation is used for treatment of PD in first human. DBS therapy to treat motor movement disorder gains FDA approval. Medtronic gains CE mark for DBS device for treating PD.	Approval for bilateral stimulation of GPi and STN for adjunctive therapy in levodopa-responsive PD. DBS for dystonia receives CE Mark and interim FDA approval, called human device exemption (HDE).	The efficacy of PPN stimulation is demnstrated using MPTP induced primates.

2005 - Date	Medtronic introduces the first rechargeable DBS device, called Activa RC. Pacemaker lifespan improved from 3 - 5 years to about 9 years.	First acute trial of on-demand DBS. A non-human primate was used.	Boston Scientific receives CE Mark for their Vercise^{TM} DBS system. The first multiple independent current source system.	–
Possible future advances (10 – 20 years)	DBS approved as a therapy for neuropsychiatric disorder such as medication resistant cases of depression, OCD and TS.	FDA approval for cranially mounted DBS. This could solve the two stage implantation procedure consisting of two separate procedures for lead and pacemaker placement.	Long distance monitoring and adjustment of patient pacemakers, which could reduce the number of face-to-face visits.	Possibility of CE Mark and/or FDA approval for first fully implantable adaptive DBS system.
	Newer DBS stimulation sites approved.	Development of devices with complete compatability for use with high Tesla MRI for better localisation of targets and improved device safety. Current devices are conditionally safe for MRI.	Robot assisted implantation through frameless stereotaxy.	

2.6 Chapter Summary

There have been various advances in DBS, most of which improve the clinical outcomes of DBS. As discussed, several factors affect the clinical outcomes of DBS, including lead positioning, programming technique, feedback signals and algorithms for closing the loop, recording devices and so on. There are currently two main methods by which DBS clinical outcomes can be improved: closed-loop DBS and directional steering in DBS. This chapter has provided some background information and a comprehensive literature on the various methods for adapting DBS. The discussions in this chapter can be summarised in the following conclusions:

a) Closed-loop DBS is a form of adaptation in DBS concerned with time-related changes. To overcome some of the limitations of DBS therapy like stimulation induced side effects and limitation of pacemaker battery life, closed loop systems would allow real-time adjustment of therapy according to quantifiable brain response. This will reduce the frequency of clinical interventions required. Potentially, with closed-loop DBS, patients are likely to benefit from fewer battery replacement surgeries [24] which can improve patients quality of life. DBS costs have been reported to be largely dominated by battery replacement cost [138].

b) Aside from adaptation through closed-loop DBS, spatial targets can be adapted and this has been found to improve therapeutic benefits.

c) Implementation of closed-loop DBS has mainly focussed on computational models, with very little breakthrough made in the transition to clinical trials.

d) The major impediment in the development of a fully implantable closed-loop DBS systems has been poor understanding of the underlying workings of DBS. Lack of insight has prevented the selection of suitable biomarkers for closed-loop DBS and has hindered the adoption of an optimum tracking and control mechanism for the disease. Currently, most experimental studies have relied on bang-bang control techniques.

e) The ability to deploy powerful feedback algorithms is limited by the allowable power consumption of fully implantable processors.

f) Promising techniques that adapt stimulation (and sensing) to targeted neural structures can be implemented using directional steering.

g) With all the research and theories on possible benefits of DBS for various disorders, there is only evidence and approval to back its use in three conditions: PD, dystonia and tremor [41]. The primary brain targets are also constrained to three: STN, GPi and Vim.

h) Finally, the future of research in aDBS is dependent on major breakthroughs in synergistic and cross-disciplinary collaborations between clinicians, engineers, statisticians, computer scientists, health care professionals, regulatory experts and most

importantly, end-users. This is the only way a clear understanding of the appropriate strategies needed to incorporate aDBS systems in patients will be achieved.

Chapter 3

Measures and Metrics for Evaluating Detection and Classification Algorithms in Brain Machine Interfaces

Research in BMI has come a long way, so far there have been six international BMI meetings, in which the most recent took place in 2016. As the field progresses, there is a need to provide objective methods to compare BMI technologies. It is for this reason studies like [143] have attempted to provide *experimental frameworks* on how to compare BMI. However, a very conspicuous gap has been the lack of standard metrics for assessing BMI systems. Drawing from Chapter 2, the major challenges of physiological signal analysis (which is central to BMI and closed-loop DBS devices), are; feedback signals, sensing techniques and feedback algorithms. This chapter is designed to delve deeper into these challenges; without limiting the analysis to only closed-loop DBS, but BMIs in general. The central theme is still closed-loop DBS, which is an instance of BMI. Thus, in a broader sense, most (or all) of the techniques applicable for use in closed-loop DBS are suitable for other BMIs, but the reverse may not be applicable. This prompted generalising and extending the review and analysis to BMI as a whole. Since closed-loop DBS is a subset of BMI. The chapter is organised as follows. The first section gives a brief context on BMI. The second section focusses on sensing techniques and feedback signals that are obtainable from the brain (external signals were not included). The third details measures that could be used to evaluate efficacy, complexity and efficiency of feedback algorithms; as well as their strengths and weaknesses. The fourth section stresses the benefits as well as limitations of standardising and customising performance metrics for feedback algorithms in BMI, such that algorithms could be assessed in a fairer, more balanced and objective way. The final section provides a summary of the salient points.

3.1 Brain Machine Interfaces (BMI)

A BMI transduces brain signals to enable some form of communication between the brain and a machine – both can take up the role of the transmitter and receiver. Figure 3.1 depicts the bidirectional communication of an idealised BMI system. There are

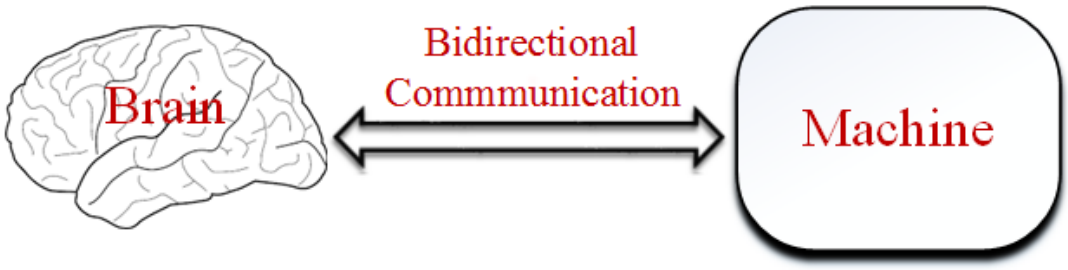


Figure 3.1: Idealized bi-directional brain-machine interface.

various ways BMIs can be classified: based on function, level of invasiveness, the origin of neural signal and by design [144]. However, more focus will be on their classification based on function. When classified based on functions, they can be broadly divided into two categories [145]. The first category are devices that relay sensory information by stimulating relevant regions of the brain concerned with various forms of perception like sound or vision to mimic their neurological function. The second category decodes neural activity in real-time to control prosthetics, motor disorders, pain and other impairment or disabilities. Among the two groups, the former controls perception to the brain based on sensory information, while the latter controls actuation to a body part or prosthetics based on recorded neural activity from the brain. Both categories can use electrical signals to relay sensory and actuation information respectively [144]. They can also be used to restore, reinforce and facilitate human sensory-motor functions [146]. Advances in neuroscience have led to a surge in BMI research. However, the major challenge remains translating research into real-life applications. These transition has mainly been hindered by limitations in sensing technology; unreliable algorithms for signal analysis and interpretation; and ineffective control strategies [147]. To address some of the challenges, proper understanding of BMI sensing technology is required as well as metrics that can be used to evaluate feedback algorithms reliably. These are highlighted in the following sections.

3.2 Sensing in BMI

In BMI, the sensing stage quantifies brain activity such that it can be used to control a prosthetic limb or to mitigate neuro-motor disorders. For effective neural recording, cutting-edge techniques that access deep and distant regions of the brain are required [148]. This has led to an upward trend in the development of miniaturised recording devices with high spatiotemporal resolution. A notable example of this is the encapsulated neural acquisition chip that records electrophysiological activities from the gyri and sulci of the brain [148]. It provides a high spatiotemporal resolution, which could lead to having more insights into brain dynamics. Major breakthrough in understanding neurophysiological dynamics is dependent on advances in BMI data acquisition [149], because the first requirement towards achieving an efficient BMI detection algorithm is acquiring neural signals without compromising their quality. This section focuses on techniques that have been used in signal acquisition for BMI appli-

cations. The list is not exhaustive, nevertheless it presents the most prominent signal acquisition techniques in BMI applications. The major focus is on actuation assistive BMIs which closed-loop DBS can be categorised into. Nonetheless, background information is provided on perception assistive BMIs.

3.2.1 Perception Assistive Sensing

These are signals that are used to help enhance sensory functions in patients with auditory, visual or any other sensory impairments. Stimulation is triggered using mainly signals external to the brain. This class consists of visual and auditory prosthesis like cochlear and retinal implants. Cochlear implants work by converting sound into patterns of electrical stimuli that are delivered using a collection of implanted microelectrodes to the auditory nerve fibre lying on the basilar membrane of the cochlea [145]. By substituting the percepts with sound and the neural system with the auditory nerve, Figure 3.2 summarises the processing chain of a cochlear implant; which is a typical example of a BMI that enhances percepts. Sound enhancing prosthesis were the first successful prosthesis. They were developed as early as the 19th century by Miller Hutchinson as hearing aids [150]. Making them the earliest BMIs to be commercially available. In 2013, the FDA approved the first retinal implant [151]. Like the cochlear implant, the retinal implant uses decoded captured images as control signals, which are then used to stimulate the optical nerve through a set of electrodes [152]. The retinal implant uses a similar processing chain to the cochlear implants as shown in Figure 3.2. With the retinal implants, the percept is vision and the neural system to be modulated is the optical nerve. Work on retinal implants is still at its early stages, and they have shown a lot of promise; however, they are still grappling with low resolution, making it difficult for blind patients to use them for daily living activities [153].

3.2.2 Actuation Assistive Sensing

These consists of the group of neural signals used to rectify, restore and ameliorate external bodily functions, mainly motor disabilities. Figure 3.3 depicts a BMI using brain signals to control a prosthetic limb. Prosthetic limbs like this restore grasp and gait abilities to seriously disabled patients, who may be paralysed or those with severe neuro-motor limitations. In restoring bodily functions, more neural information can be obtained using innovative approaches having high spatiotemporal resolution or techniques measuring multiple brain activities [149]. Regarding neural activity measured, this review broadly classifies them into two categories: electrical and metabolic activity. Below is a brief description of the various modalities used in obtaining electrical and metabolic activity from the brain.

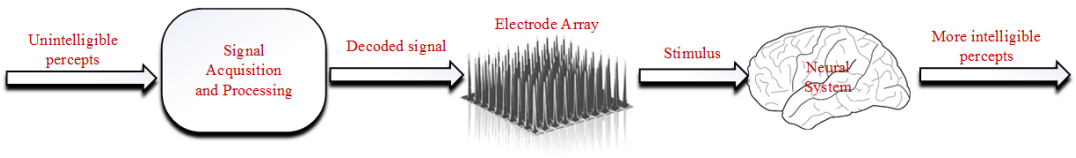


Figure 3.2: Processing chain of perception assistive BMI.

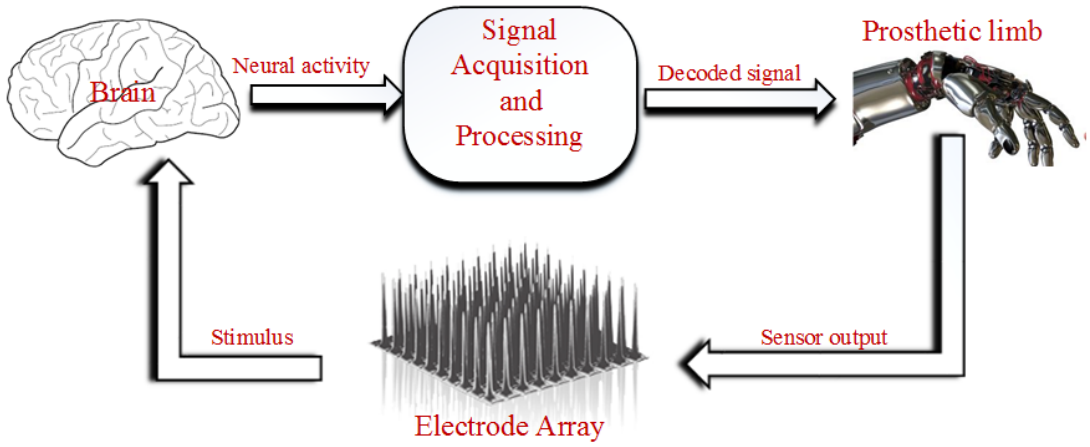


Figure 3.3: Actuation assistive BMI using brain signals to control a prosthetic limb.

3.2.2.1 Electrical Activity Sensing

Neuro-electrophysiology has been used in studying bio-electrical properties of brain cells and tissues. Notable among, was the revolutionary discovery by Hodgkin and Huxley in 1952, on the initiation of action potentials in squid axons which eventually led to a Nobel Prize [154]. Later, there was the discovery by Hubel and Wisel in 1977, about how individual neurons contribute to visual processing [155]. These ground breaking studies set the pace in neuro-electrophysiology. To extract useful information from neuro-electrophysiological signals, a good understanding of how these signals are formed at the neural level is required. Generally, neuro-electrophysiological signals represent the spiking behaviour of a single neuron, a small neural ensemble and the mean potentials of a large neural ensemble. The larger the neural population, the higher the amplitudes as more neurons contribute additively to the signal. The structure of the neuron was discussed in Chapter 2. Neuro-electrophysiological signals are used for various applications in clinical settings. Primarily, they have been used in BMI, where implanted devices are used to record and decode brain signals, which are used in controlling external machines, like prosthetic limbs [57]. Additionally, electrophysiological signals are used in localising areas where seizures begin in both medically tractable and intractable epilepsy [156]. They have been proven to be good markers for movement disorders such as PD, ET and dystonia [16]. Their use is extending to tracing neuropsychiatric disorders like OCD, dementia, attention deficit hyperactivity disorder (ADHD), Alzheimer's disease (AD), and schizophrenia (SZ) among others [157]. Fundamentally, feedback signals should be selected based on their: level of invasiveness, resolution, signal content, and clinical relevance. The selection of signals will ultimately depend on the design of the entire system in relation to the signal processing capability available. The following sections give a brief description of the sensing techniques and signals obtainable by electrical activity sensing in the brain.

Neuro-electrophysiological Recordings

Intracellular recording is the measurement of voltage or current within the membrane of

a cell. This is done by inserting an electrode in the cell and a reference electrode outside the cell. This could be done using a current or voltage clamp [158]. In a current clamp, current is injected through the intracellular electrodes and the resulting membrane potential is measured using an amplifier. Whereas, the voltage clamp holds the membrane potential at a fixed value and the current flowing through the intracellular electrode is measured. The major techniques used in intracellular recording are measurements of current, potential and conductance. On the other hand, extracellular recording is the main method for measuring in vivo neural activity. For a single neuron, extracellular recording is achieved by placing an electrode close to the neuronal soma such that the firing rate of the neuron is measured by the number of spikes [159]. Extracellular recording has been more prevalent due to its ability to provide neuronal activity, coupled with its relative ease of use compared to intracellular activity. Aside from single neuron activity, research is growing in the study of how a network of neurons influences various functions like cognition, movement and perception. These studies have mainly used extracellular recordings using multi-electrode arrays (MEA). Extracellular potentials provide information consisting of high-frequency spiking activity (> 500 Hz), which stem from some neurons within the immediate vicinity of recording electrode and are termed multi-unit activity (MUA). And the low-frequency potentials consisting of LFP. Figure 3.4 presents the conventional set-up for measuring extracellular activity from a neural population [160]. An ideal measurement technique is required to be able to provide activity of single neurons, at the same time providing whole brain activity, within a microsecond time scale [159]; which might only be achievable by combining recordings from various techniques.

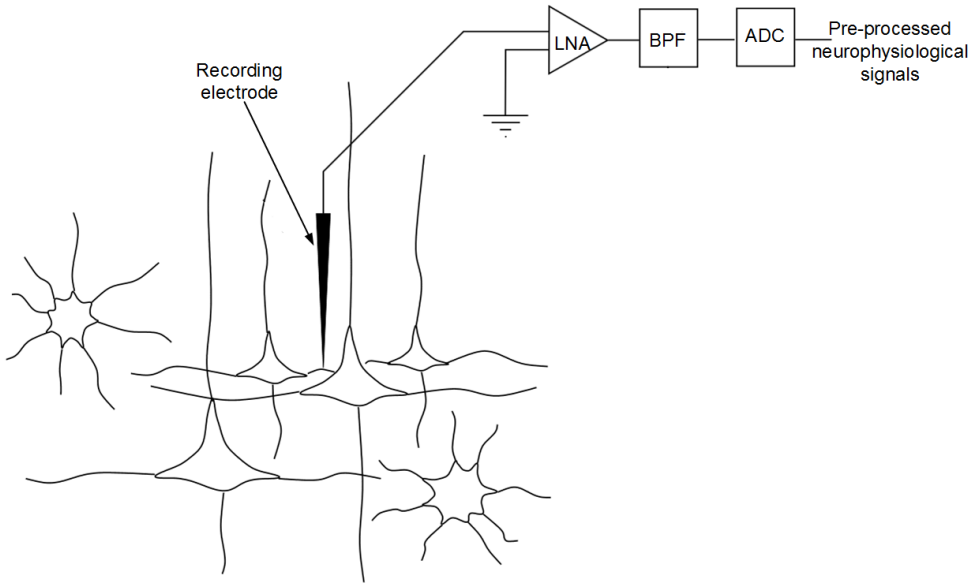


Figure 3.4: The basic set-up for measuring and analysing extracellular neural signals. The set-up shows the low noise amplifier (LNA), band pass filter (BPF) and analog to digital converter (ADC).

Unit Activity

Using sharp extracellular electrodes as in Figure 3.4, action potentials are generally extracted from a single neighbouring neuron (single-unit recording) or from an unknown population of neighbouring neurons (multi-unit recording) [161]. These are mostly high-frequency extracellular potentials ($>$ 500 Hz). Single unit activity is used in understanding how a neuron responds to specific stimulus or to understand correlation between various neurons. This has led to their use in providing insight into patterned activity within the STN and GPi in relation to movement, cognitive processes and memory; making them potentially suited as biomarkers for use in closed-loop DBS. Nevertheless, they are hindered by recalibration (due to drift in neuronal properties), need for precision on target neuron and unreliability of recording over extended use [16]. Single neuron recordings represent certain movement features; nonetheless, they have an increased sampling rate requirements, degradation at the neuron-electrode interface and difficulty maintaining recordings from the same neuron for extended periods of time. However, the difficulty in maintaining recordings has been as a result of the size of recording electrodes, which mostly picks up neuronal ensemble activity as against the required single unit activity [162]. This has necessitated the need for additional processing, which is spike sorting. It is used to extract single unit activities from multiunit activities. Compared to other neuro-electrophysiological signals, unit activities are useful in BMI applications, since high spatial resolution is required. Spikes have been found to show a clear relationship to movement and behavioural functions as highlighted by their application in BMI for prosthetic limbs. This has led to their use as biomarkers for regulating stimulation in closed-loop DBS [75].

Local Field Potentials

LFP are low frequency ($<$ 500 Hz) extracellular potentials obtained from time sampling a localised population of neurons. LFPs can simply be measured with a standard EEG amplier connected to the implanted DBS electrode, and are generated by summated postsynaptic potentials resulting from excitation in basal ganglia and cortical neurons [163]. Unlike unit activity, they tend not to drift over time which makes them more reliable and stable [162]. Because of their population based nature, they are more informative, due to their time and frequency response; and offer a better trade-off between high spatial resolution (common in unit activity) and high spatial scale (common in global field potentials) than other neuro-electrophysiological signals. Spatial scale deals with the level of spatial coverage of the measuring technique, that is, on a macro level. On the other hand, spatial resolution deals with the level of detail in spatial information that can be obtained, that is, on a micro level.

Another advantage of localised population of neurons is the long-term experience in signal processing for EEG-like signals researchers have, particularly LFPs. Figure 3.5 illustrates the region where LFP recording can be obtained [9]. Also, LFP processing algorithms can be easily implemented on microchips, and are therefore ideal for use in implantable devices. Since LFP processing microchips have been adopted in several studies, using LFP means no additional work or procedures to be undertaken. This

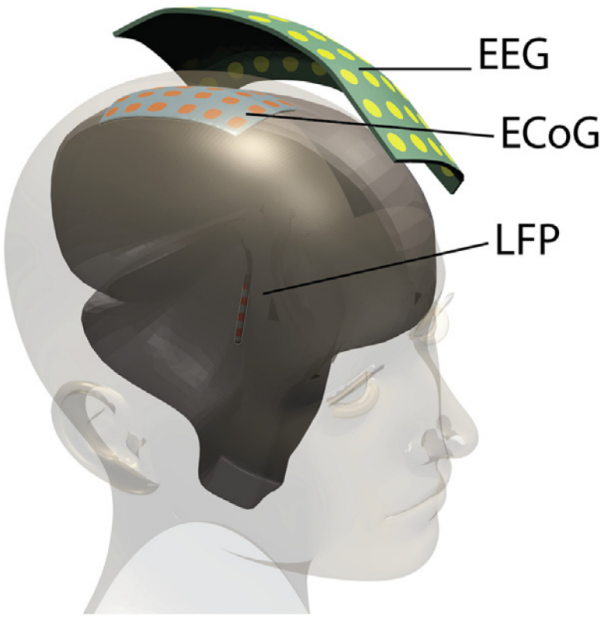


Figure 3.5: Candidate neuro-electrophysiological signals and their recording sites (adapted from [9]).

makes them ideally suited for many applications requiring neural signals for feedback. As biomarkers for closed-loop DBS, current evidence supports the hypothesis that LFP activity changes in response to the patient's clinical state. Basal ganglia LFPs oscillate in several frequency bands, ranging from; very-low frequencies ($3-8$ Hz), alpha ($8-12$ Hz), beta frequencies ($13-30$ Hz), gamma (>30 Hz), and very-high frequencies ($250-350$ Hz). The most studied and discussed LFP oscillations are beta frequencies because they seem to reflect the patient's motor state. Changes in beta LFP activity mainly reflect basal ganglia responses to dopamine and correlate with motor performance.
Even though the exact mechanisms of DBS are still under debate, ample evidence shows that LFP oscillations in patients with PD are specically modulated by DBS [164]. Notwithstanding, LFPs have their limitations, some of which are listed below:

- There is evidence that LFPs correlate closely with the individual patient's motor status, but correlation across patients is yet to be established [16].
- Recent studies suggest that other activities like cognitive and behavioural functions might modulate LFP [164].
- Abnormal oscillations reflect some clinical impairment in patients, but direct relationships are yet to be established [60].
- Conversely, there is no correlation between beta band LFP and UPDRS score, making some studies to downplay its importance [165].

LFP has a relatively better correlation with PD compared to other neuroelectrophysiological signals. Various clinical studies have employed it as biomarker for PD [24,58,87] and numerous computational studies have been concerned with its manipulation in order to stabilise patients state [61,68,69,80,81,105].

Global Field Potential

These are activities from a much larger population of neurons than LFP. Like LFPs, EEG measures the summed electrical activity of many neurons and is measured with electrodes at the surface of the scalp; though at a larger scale. Figure 3.5 depicts the recording sites for EEG and electrocorticograhy (ECoG) which are examples of global field potentials. From Figure 3.5, it can be seen that EEG are subject to filtering due to propagation through various media such as the cranium, dura mater, cerebrospinal fluid and other surrounding tissue [159]. Due to this frequency filtering, action potentials are severely attenuated and are not visible on EEG electrodes. Also, low-frequency activities such as synaptic potentials dominate EEG signals, since they can propagate over large distances within the extracellular space. EEG recordings provide average activities of neurons on the order of $10^5 - 10^8$ [166], this inhibits both their spatial and temporal resolution. On the other hand, ECoG can be used to overcome some of the shortcomings of EEG. ECoG measurements are made on the surface of the cortex as is shown in Figure 3.5, which results in less filtering compared to EEG which is made on the scalp. Like EEG, ECoG measures the summated electrical activity of many neurons, to be precise, it uses the summation of between $10^2 - 10^3$ neurons [166]. This enables ECoG to record frequencies up to 200 Hz compared to the less than 70 Hz present in EEG [167]. Thus, EEG and ECoG make the same measurements, though EEG signals have poorer spatiotemporal resolution due to filtering by cortical tissue and their large coverage area. In BMI, higher frequency contents in signals mostly lead to more relevant information for decoding patient state and intentions; which are necessary for BMI control [168]. With the lower frequency-specific information content of EEG compared to other neuro-electrophysiological signals, their long-term use as biomarkers may be unsuitable. Nevertheless, they relatively have a higher spatial scale compared to the other signals since they cover the whole brain; making them suitable for acquiring more general fluctuations in frequency information across the brain. This property makes them ideally suitable as complementary signals in BMI applications. As a result of this, they have found application in responsive stimulation for epilepsy. Their adoption in epilepsy makes them an option for PD applications since epilepsy is a neurological disorder impairing movement and other behavioural functions just like PD.

3.2.2.2 Metabolic Activity Sensing

For PD and DBS, hemodynamic or neuro-transmitter response could be used in obtaining vital brain information. In DBS, excitation and inhibition of neuronal signals occur as a result of stimulation and improvement in conditions. In addition, other secondary effects are blood flow changes, modulation of neurotransmitters, neurogenesis and a host of other metabolic activities [27]. This makes the investigation of metabolic activity for closed-loop DBS important. In the hemodynamic response, blood releases glucose to active neurons at a higher rate than in the area of inactive neurons [169]. The glucose and oxygen released into the blood stream results in an increase in oxyhemoglobin in the veins around the active region. Hemodynamic changes in PD patients can serve

as good biomarkers since DBS and PD induce cortical hemodynamic changes in patients [170]. These changes can be detected by methods such as functional magnetic resonance imaging (fMRI), diffusion magnetic resonance imaging (dMRI) and near-infrared spectroscopy (NIRS). Like in hemodynamic responses, the use of techniques that measure neuro-transmitter response is pertinent because PD results in degeneration of cells that use dopamine as neurotransmitters [3]. Monitoring dopamine traces from cerebral metabolites have been reported [16], but miniaturisation of chemical analysis is a major barrier. In addition to monitoring neurotransmitter and hemodynamic responses, responses to specific molecules are measured using optical micro-imaging techniques mostly by using fluorescence measurements. Recent advances in optical imaging techniques have led to single-cell resolution in functional neuroimaging which uses a two-photon microscope [171]. Understanding pathological brain processes down to the single neuron level is necessary towards harnessing the ability of closed-loop DBS as well as BMIs to restore bodily functions. The advantage optical approaches have over other methods is that they have high spatial localisation and are relatively less invasive ($\sim$ 1 - 2 mm in depth) compared to other methods [172]. However, their major shortcoming is their low temporal resolution compared to neuro-electrophysiological methods. This is partly due to their high requirement for signal processing and data analysis [173]. Below is a summary of the major methods used in analysing and obtaining metabolic activity from the brain that can be used in closed-loop DBS.

Florescence Measurements

Fluorescence measurements have a distinctive response to the presence of specific molecules like sodium, potassium or calcium. Fluorescence measurements of neuronal activity can be classified into two categories: those that are sensitive towards membrane voltage and those that detect changes in intracellular calcium concentration [173]. Sensors sensitive to membrane potentials produce relatively small signals in response to action potentials. Currently, calcium-sensitive sensors are orders of magnitude more sensitive than sodium or potassium sensitive sensors. Since, in principle, initiation and propagation of action potential can give rise to about a hundred times higher calcium concentration that under rest [174]. This can be used to measure active and inactive neurons in the brain.

Near-Infrared Spectroscopy (NIRS)

NIRS uses optical spectroscopy based on infrared light to quantify changes in cerebral metabolism due to neural activity. Infrared light penetrates the human skull up to a depth of about 1- 3 cm below the skull. This enables NIRS to measure concentration of oxyhemoglobin based on light attenuation (absorption and scattering) [175]. The shallow penetration of light makes it not very ideal for applications like PD that monitor biomarkers in deep regions of the brain. Nevertheless, it is a promising diagnostic tool to investigate neurovascular coupling, for example in epilepsy to develop novel early seizure detection algorithms, because vascular changes occur about 100 milliseconds after the associated neural activity. Which is an acceptable temporal resolution for both

closed-loop DBS and BMI applications [176]. In addition, it has a spatial resolution of within 1 cm. The major advantage of adopting optical modalities like NIRS and fluorescence measurements in recording neural activity is their high specificity, which will go a long way in facilitating artefact free closed-loop DBS systems. NIRS was first used in DBS patients by Sakatani and colleagues [177]. The finding suggested that therapeutic benefits were reflected by changes in oxyhemoglobin levels in the prefrontal cortex. Despite its bulkiness, NIRS has been proposed as a suitable measure of neuronal activity due to its ability to accurately quantify neuronal activity which is reflective of symptom severity and has been proposed as a candidate signal to adjust the parameters of DBS in a closed loop configuration [178]. Applications utilising this technology for BMIs and closed-loop DBS are still in infancy. Thorough studies using a large number of patients that produce encouraging results are required so as to establish its feasibility.

Magnetic Resonance Imaging (MRI)

MRI is an emerging technology for observing neural activity in the living brain. It has tremendous potential for use in applications like blood-oxygen-level-dependent (BOLD) fMRI, which is a non-invasive method for monitoring brain functions [149]. Like NIRS, fMRI is a measurement based on hemodynamic changes, and it offers a spatial resolution in the millimetres range. It has been shown to offer tremendous insights into the underlying dynamics of the human brain [179]. Also, understanding the underlying mechanisms can give more insight as to why different patients' brains respond differently to similar levels of stimulation. Aside from fMRI, the complex activity in the white matter of the brain can be captured in great detail using dMRI [180]. It captures the molecular displacement of water within a voxel. This is used to deduce the location and direction of white matter tracts based on the directional diffusion of water. Changes in white matter fibre tract have been used to understand the internal working of the brain. Muetzel et al. [181], confirmed that changes in white matter fibre tract connection correlated with subject performance in specific tasks in healthy patients.

Fast Scan Cyclic Voltammetry (FSCV)

FSCV is a voltammetry technique that applies a linearly varying potential through carbon fibre microelectrode (CFM), resulting in redox chemical reactions around the electroactive molecules [182]. The concentration of analytes is measured by the magnitude of evoked current peaks to the redox reaction at the electrode surface. The relationship between the applied voltage versus the resulting current provides a chemical signature for the presence of certain neurotransmitters or analytes. FSCV detection is mainly limited to electroactive analytes; electroactive molecules like dopamine (a biomarker for PD), adenosine (a biomarker for sleep), and oxygen (which signifies the presence of anoxic brain injuries). The major limitations of FSCV are its bulkiness and that the lifetime of CFM is a few months, which restricts the application of FSCV detection to intraoperative approaches. For DBS, using an anesthetised rat model, the MINCS in [103], was interfaced to FSCV to wirelessly regulate stimulation as a proof-of-principle test for closed-loop DBS using neurochemical signals for feedback.

Intracranial Micro-dialysis

Microdialysis is the most commonly used method to measure the chemical concentration of analytes in the brain [183]. It uses a dialysis probe that penetrates very small molecules in the brain. The brain is supplied with artificial cerebrospinal fluid, the quantity of molecules that diffuse into the probe and the dialysate are collected and analysed off-line. Its spatiotemporal resolution is not excellent because a certain amount of dialysate has to be collected before any analysis can be done; which impedes time resolution. However, it has a very high degree of chemical selectivity and sensitivity. Due to its poor temporal resolution, it can only be used to measure long-term changes in analytes or neurotransmitter for use in closed-loop or BMI applications. Its sensitivity and selectivity make it suitable for applications like home-based monitoring of PD patients. This could go a long way in reducing the frequency of face-to-face visits. Table 3.1 summarises some of the characteristics of various feedback signals that are suitable for closed-loop DBS.

Table 3.1: Comparison of possible feedback signals and sensing techniques for closed-loop DBS.

Electrical Activity Measurements	Spike	LFP	ECoG/iEEG	EEG	
Activity Measured	Unit activity	Average potential of a localised neural population	cortical or intracranial activity	Electrical activity from scalp	
Temporal Resolution	less than 1 ms	$\sim$ 1 ms	$\sim$ 3 ms	$\sim$ 50 ms	
Spatial Resolution	$\sim$ 50 μ m	$\sim$ 0.5 mm	$\sim$ 1 mm	$\sim$ 10 mm	
Level of Invasiveness	Invasive	Invasive	Minimally invasive	Non-invasive	
Practicability for use in closed-loop DBS	Implantable	Implantable	Implantable	Bulky, but cheap	
Metabolic Activity Measurements	**Fluorescence Measurements**	**NIRS**	**fMRI**	**FSCV**	**Intracranial dialysis**
Activity Measured	$Ca^{2+}, Na^{+} or K^{+}$ concentration in the brain	Concentration of oxyhemoglobin	blood-oxygen-level and molecular displacement of water	Concentration of neurotransmitter (dopamine)	Concentration of neurotransmitter (dopamine)
Temporal Resolution	$\sim$ 30 ms	less than 1 s	$\sim$ 1 s	$\sim$ 1 s	$\sim$ 1 s
Spatial Resolution	$\sim$ 10 μ m	$\sim$ 5 mm	$\sim$ 1 mm	$\sim$ 30 μ m	$\sim$ 200 μ m
Level of Invasiveness	Invasive	Non-invasive	Non-invasive	Invasive	Invasive
Practicability for use in closed-loop DBS	Implantable (with very high data analysis cost)	Bulky	Bulky (a major hindrance is DBS devices are still MR conditional)	Bulky (with high specificity and selectivity)	Bulky (with high specificity and selectivity)

3.2.2.3 Electrical versus Metabolic Activity

For effective neural recording, cutting-edge techniques that access deep and distant regions of the brain are required [148]. These could lead to more insight in brain dynamics. Nonetheless, of equal importance are techniques that have spatial coverage. A breakthrough in understanding neurophysiological dynamics is dependent on advances in neural signal acquisition [149]. This is the first requirement towards achieving an efficient closed-loop DBS system. Changes in the bio-chemical environment within the brain can be representative of intended actions and actual actions in patients. These characteristics make metabolic activity sensing suitable for quantifying neural activity. The use of metabolic activity as biomarkers have been investigated in [184]. Their major shortcomings are safety concerns like MRI compliance and metal artefacts. However, some other metabolic activity sensors such as NIRS are not affected by metal artefacts, but have relatively poor temporal resolution compared to electrophysiological activity. Their large size is a stumbling block towards attaining fully implantable closed-loop DBS systems. Apart from sensitivity to metal artefacts in fMRI, metabolic activity offers many advantages compared to electrical activity recording, notably: absence of electrical noise, simultaneous imaging of a large number of neurons and selective recording from genetically-targeted regions of the brain [185]. Their high SNR, specificity and selectivity can go a long way towards facilitating artefact free closed-loop DBS systems. For electrical activity, information content is dependent on spatio-temporal resolution, with EEG and single unit activity on the extremes of the spectrum: EEG has the highest spatial scale and the least temporal resolution, while single unit activities have the highest temporal resolution and the least spatial coverage. LFPs offer a compromise in terms spatiotemporal resolution. Combined with their long-term stability at the ETI, this makes them very attractive feedback signals for closed-loop forms of DBS [16]. However, the pertinent question is how informative they are compared to other neuro-electrophysiological signals?

Closed-loop DBS applications can adopt effective paradigms that combine both metabolic and electrical activity sensors for acquiring brain responses in real-time which increases spatiotemporal resolution. This leads to better identification of disease and non-disease states in patients, as the level of information content is strictly dependent on spatiotemporal capabilities of the sensor. This complementary approach takes advantage of the best of both worlds: the fast response rate of electrical activity, and the slower more specific, selective and qualitative sensing offered by metabolic activity. The complementarity of sensing electrical and metabolic activity have found application in BMI [149,186].

3.3 Measures for Evaluating BMI Detection (Feedback) Algorithms

With the advancement in the use of BMI for neural rehabilitation, the rise in the significance of detection algorithms as their major building blocks cannot be overemphasised. In closed-loop DBS or BMI applications, detection algorithms convert recorded brain activity into useful information that could be translated or used as the control signals

for the control of sensorimotor functions. Detection algorithms are the first step towards developing BMIs that uses pathological brain activity to ameliorate, mitigate or restore bodily function in patients with disabilities such as such as PD, Alzheimer's disease, epilepsy, spinal cord injury (SCI) and so on. They are the major cornerstone of the BRAIN initiative (Brain Research through Advancing Innovative Neuro-technologies), which is a $100 million investment in research and development with the goal of supporting the development and application of innovative technologies that can create a dynamic understanding of brain function [187]. Efficient and efficacious detection algorithms could lead to fully implantable BMI systems with real-time capabilities that could be used to monitor disease progression in patients. Also the economic impact is another motivation for understanding the various metrics that can be used in evaluating detection algorithms. Currently, there is an increase in the cost of management for neuro-motor disabilities both for individuals [188] and health care administrators [189]. As such, it is paramount to review and analyse various performance metrics in BMI detection algorithm as they influence this cost of BMI detection algorithms. This will enable proper resource utilisation for hardware efficient implementation. The focus will be only BMI detection algorithms. The metrics were categorised into three, based on what characteristic they evaluate. The three categories are complexity, efficacy and efficiency. Of these three categories of measures, there are more subtle differences between efficacy and efficiency. Efficacy means the ability to produce the desired results irrespective of the resources expended. On the other hand, efficiency is not only concerned with achieving results but also concerned with the resources utilised in achieving the results.

3.3.1 Efficacy Measures

We review various methods that are currently used in assessing the efficacy of detection algorithms in BMI. Evaluating the efficacy of detection algorithms can be tedious, as there are several nearly similar measures with rather subtle differences that can be used. However, some methods are more effective for some BMI applications based on the peculiar features of the said application. While others perform better at revealing certain characteristics of the detection algorithms that may ordinarily be less conspicuous. More emphasis is placed on metrics that are good at evaluating binary decision problems, like disease and non-disease states. A potential pitfall in analysing binary detection algorithms is data redundancy [90]. That is, using a similar sequence of examples to train and test algorithms could lead to having an overestimated performance because it only ends up producing the output to a possible training input rather than interpolating or extrapolating, which is the main aim of a detection algorithm [190]. At its most basic level, the efficacy of binary detection can be assessed using the confusion matrix. Other efficacy measures are: difference and information measures. This review will look at several metrics for assessing efficacy, with the intention of deciding which metrics are ideally tailored for our application in subsequent chapters.

3.3.1.1 Confusion Matrix

The confusion matrix or contingency table is a table that contains information on the actual and detected states for a supervised learning classifier. For a binary classifier it is a 2×2 matrix consisting of the cells in Table 3.2. However, when the detection problem is an M-ary detection problem, then the confusion matrix can be represented by an $M \times M$ matrix. From the confusion matrix in Table 3.2, the various quantities are defined as follows: TP represents the true positives, FP represents the false positives, FN represents the false negatives and TN represents the true negatives. Also, PD^+ occurs when an epoch is detected to have a PD state, while PD^- are PD-free segments. However, using a range of numbers to represent performance as in the confusion matrix, it may not be obvious how well a detection algorithm performs. An alternative will be to use single performance measures based on a combination of two or more relevant quantities from the confusion matrix. There is no single measure that gives a completely fair assessment of the confusion matrix. Nevertheless, there are measures that give a very good picture such as the F1-score and Mathews correlation coefficient (MCC). They give a very balanced measure for uneven classes. Below is a brief description of the various performance metrics obtainable from the confusion matrix that have been used in various literature on BMI.

Table 3.2: Confusion matrix summarising performance metrics [10].

		Actual State PD^+	Actual State PD^-	
Detected State	PD^+	TP	FP (type I error)	Precision or Positive Predictive Value $= \frac{TP}{TP+FP}$
	PD^-	FN (type II error)	TN	Negative Predictive Value $= \frac{TN}{TN+FN}$
		Sensitivity or Recall or True Positive Rate $= \frac{TP}{TP+FN}$ False negative rate or Miss rate $= \frac{FN}{TP+FN}$	Specificity or True Negative Rate $= \frac{TN}{TN+FP}$ False positive rate or Fall out $= \frac{FP}{TN+FP}$	Classification accuracy $= \frac{TN+TP}{TN+FN+TP+FP}$

1) Classification Accuracy and Classification Error: Classification accuracy measures the proportion of correctly classified (detected) states. On the other hand, classification error measures the proportion of classifications that were wrong. Both classification accuracy and error are not suitable metrics for measuring classifier quality in applications with skewed classes. To overcome this, newer metrics like the weighted classification error (WCE), which applies various weights to errors in the different classes depending on which error is more likely to occur and which is more (or less)

desirable [191]. WCE can be represented mathematically as,

$$Error_{weighted} = (\frac{\alpha \cdot FP}{FP + TN}) + (\frac{(1-\alpha) \cdot FN}{TP + FN}) \tag{3.1}$$

where α is the weight (or penalty) applied to the error in the non-PD class and $1 - \alpha$ is the weight applied to the error in the PD class. The error in the PD class represents type I error (false-positive rate), while the error in the non-PD class represents type II error (false-negative rate).

2) Sensitivity and Specificity: The sensitivity or recall of a detection test gives the percentage of patients for whom the outcome is positive that are correctly detected by the test – true positive rate. On the other hand, the specificity gives the percentage of patients for whom the outcome is negative that are correctly detected by the test – true negative rate.

3) Positive and Negative Predictive Value: Precision or positive predictive value (PPV) of a test is the probability that a patient has a positive outcome given that a positive test result is obtained. This is different from sensitivity, which is the probability that a patient has a positive test result given that a positive outcome obtained. Similarly, the negative predictive value (NPV) is the probability that a patient has a negative outcome given that a negative test result is obtained. In contrast to specificity, which is the probability that a patient has a negative test result given that they have a negative outcome.

4) Miss and Fall-out Rate: The miss rate gives the conditional probability of a negative test result given that the condition being looked for is positive. The fall out rate gives the conditional probability of detection being positive given an event that it was not positive. It gives the significance level of the test. In a sense, the miss and fall-out rate are the opposites of sensitivity and specificity, respectively.

5) Choice Probability (CP): Receiver operating characteristic (ROC) is a plot of sensitivity and false positive rate and has an area under the curve (AUC) of between 0 and 1. It is used to evaluate the performance of various detection algorithms. The area under the curve (AUC) of the ROC, also called the choice probability (CP), represents the probability that the detector will correctly classify an event in a two-alternative forced-choice classification. This is described in more detail in [192].

6) F1-score: In a model consisting of skewed classes, for example $PD = 2\%$ and $non - PD = 98\%$, classification accuracy is not a good metric for analysing such models as a randomly guessing classifier that classifiers all test cases as non-PD will achieve 98% accuracy. This is obviously misleading. In such situation, a single useful metric that can be used to analyse the performance of a classifier is the F1 score, which is obtained using the precision and sensitivity (or recall). Mathematically [10],

$$F_1 = (\frac{2PR}{P+R}) \tag{3.2}$$

From the equation, the F_1 score uses the precision (P) and recall (R) to determine the performance of a classifier with the lower of the two dominating. This is because high precision and high recall are required for PD detection to properly modulate therapy. High precision is required because wrong prediction will result in administering stimulation when it is not required, and this may lead to side effects [36]. On the other hand, high recall is also required to avoid missing too many PD cases. As a high number of PD misses will worsen the patient state since the patient will be starved of the required stimulation needed to mitigate the effects of PD [37].

7) Mathews Correlation Coefficient (MCC): Another balanced measure for disproportionate classes is the MCC. It is a modified version of the Pearson correlation coefficient. It has a range between -1 (total disagreement) and $+1$ (total agreement). It is normally 0 for totally random predictions, producing a value of 0 for completely independent variables. Mathematically [10],

$$M_{CC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}} \tag{3.3}$$

The major shortcoming of MCC is that it can only be used when one of the denominators $TP+FN$, $TP+FP$, $TN+FP$ and $TN+FN$ is not a zero [10]. This makes the F1- score a more practical measure when assessing detection algorithms for test cases with uneven classes. Table 3.3 summarises the strengths and weaknesses of various measures obtainable from the confusion matrix.

Table 3.3: Strengths and weaknesses of various measures from the confusion matrix.

Efficacy measure	Strength	Weakness	Ref.
Sensitivity, Specificity and Precision	Provides information on only one characteristic. This may come handy in applications where a single efficacy characteristic is relevant.	Seriously misleading when used to assess the overall performance of a detection algorithm.	[193]
Classification accuracy and error	Simple, straightforward and popular.	Can be very misleading for skewed classes. Weighted classification error can be used to address this.	[149]
Choice Probability and ROC	An efficient way to obtain information on the relationship between sensitivity and specificity. A fairly balanced measured for classification with skewed classes.	AUC dependent on curve fitting method. Areas without patient data can be highly extrapolated.	[194]

F1-score	A balanced measure for classification with skewed classes. Also, provides information on precision and sensitivity.	The lower of the two between sensitivity and precision dominates and influences the measure.	[195]
MCC	A balanced measure for classification with skewed classes.	It can only be used when one of the denominators $TP+FN, TP+FP, TN+FP$ and $TN+FN$ is not zero.	[88]

3.3.1.2 Difference and Information Measures

These are measures that quantify relative discrepancies between the actual results and detected results. Difference measures are performance measures that identify the number of positions (or places) at which predictions differ from the actual results. Weighted distance measures can be used in evaluating skewed classes to account for the disproportionate classes or as a custom accuracy measure in balanced classes. Information measures give an insight into the amount of information lost when a predictive model is used to approximate the actual model.Table 3.4 and Table 3.5 summarises the strengths and weaknesses of difference and information measures respectively.

Table 3.4: Strengths and weaknesses of difference measures.

Efficacy measure	Strength	Weakness	Ref.
Hamming distance	Easy to implement.	Misleading when used to compare scenarios with different number of detection instances.	[196]
L^P distances	The p term could serve as a good penalty term for applications demanding high performance.	The choice of p could overstate or understate performance.	[197]
Projection error	It measures how well the low dimensional representation captures the features in the original data.	It can be computationally intensive when data consists of a large number of observations as well as dimensions.	[198]

Table 3.5: Strengths and weaknesses of information measures.

Efficacy measure	Strength	Weakness	Ref.
Information gain	Captures probabilistic information which is important in applications with a defined probability distribution.	It is a relative measure. So it only accurately measures relative convergence to a specific model.	[199]
AIC	Not only measures efficacy, but parsimony as well.	Inadequate for assessing simple processes.	[200]
BIC	Not only measures efficacy, but parsimony as well.	Inadequate for assessing complex processes.	[201]
Cross correlation	A simple and straightforward way of comparing data sequences.	Can be misleading when used to compare performance on signals with small and large amplitudes. As it exaggerates the performance of the large amplitude signal due to the dot product. A possible solution to this is by using the normalised cross correlation.	[202]

3.3.2 Complexity Measures

Real-time detection in BMI algorithms can only be realised with low computation algorithms. To facilitate implantation, this algorithms need to be optimised such that their microchip implementations have significantly small scales in terms of size and power consumption. By leveraging on state of the art CMOS technology and computational techniques (low complexity, high speed and low-power); on-line BMI detection processors can be implemented. This could lead to efficient BMI systems that can be deployed for chronic clinical use.

3.3.2.1 Arithmetic Operators

A critical consideration in processor designs for BMI is to determine the number of operations required to decode information from neural activity. Arithmetic operations are required on recorded signals in order to transform them into useful, manageable or computationally efficient forms [9]. Detection algorithms for BMIs mainly involve three stages: feature extraction, dimensionality reduction and pattern classification. Bio-signal processing mainly involves operators that conduct time-domain processing, spectral processing and a mixture of both in some cases. These processing are mainly done using fixed-point or floating-point number representation which are outside the scope of this review. However, arithmetic operations are seen through the lenses of addition and multiplication, which form the basis for all arithmetic operations. The total number of arithmetic operations in a DSP computation is dependent on the resolution

of the digital converter used. Minimising complexity without compromising performance can be achieved by optimising the resolution of the digital converter. Also, complexity can be reduced by discarding (or reformulating) all redundant variables (or parameters) in a computation [203]. Computationally efficient algorithms are those that scale logarithmically with an increase in a dependent variable.

3.3.2.2 Hardware Resources

Detection algorithms for closed-loop DBS and BMI applications have essentially carried out off-line processing by tethering patients/participants or wirelessly transferring the data to high-performance computers. Tethering is impractical as it hinders normal patient activities. And wirelessly transferring data comes with bandwidth and power restrictions that may not be easily met. Hardware resources are determined by key parameters like power and area. For low-power implementation, there is a shift towards the use of microchip technologies that embed signal processing. Recent advancements in microchip technology have led to the development of detection algorithms for BMI applications. Several microchip implementations of closed-loop DBS systems using neural features and incorporating stimulation have been reported for various applications. Table 3.6 summarises those published between 2014 to date. All the systems are proof-of-concept implementations using mostly offline processing and were mostly used for closed-loop DBS control of epileptic seizures [114, 204–208]. From Table 3.6, it is clear that applications using signals with low spatial scale (and high temporal resolution) e.g. spikes (unit activity) employ less processing area and power than applications with much large spatial scale (and low temporal resolution) like ECoG and EEG. These confirms the view expressed [209], that event-related potentials have been shown to be more effective in BMI applications compared to spontaneous signals such as EEG because they do not require large storage requirements and lengthy training periods. The requirements for chronically implanted microchips are becoming more and more stringent to avoid neural tissue damage. This makes the need for minimising power and area in state-of-the-art microprocessors ever more important.

Table 3.6: Closed-loop DBS systems with online recording and stimulation.

Ref.	Year	On-chip Stages	Neural Signals	CMOS Tech.	Processing	Power (μW)	Area (mm^2)
[204]	2017	Recording, feature extraction and stimulation	EEG and ECoG	0.13 μm	FPGA	1 300	12
[210]	2017	Feature extraction and stimulation.	Spike and LFP	0.18 μm	CPU	896	3.7
[205]	2016	Recording and stimulation	ECoG	0.13 μm	FPGA	2 170	16
[206]	2015	Feature extraction, data compression and stimulation.	Spike	65 nm	CPU	626.4	4.8

[207]	2015	Recording and stimulation	ECoG	0.35 μm	FPGA	13 500	12.8
[208]	2014	Recording, seizure detection and stimulation.	iEEG	0.18 μm	On-site	2 800	13.5
[114]	2014	Recording, feature extraction and stimulation	Spike and LFP	0.18 μm	Partly on-site	468	4

3.3.3 Efficiency and Optimality Measures

This introduces efficiency metrics used in evaluating detection algorithms. In detection algorithms, efficiency metrics or measures provide information on the trade-off between resource utilisation and the performance of detection algorithms.

3.3.3.1 Energy Efficiency

Energy efficiency is a metric defined around operations like multiplications, additions or delays. This is the number of useful operations divided by the energy required to do them. It is measured in operations/Joules, and it gives the average number of Joules required for each operation. It is a metric that can be used to assess how optimal an algorithm is compared to other algorithms performing the same function [211].

3.3.3.2 Energy-Delay Product (EDP)

EDP gives the average energy consumed per switching event. In this context, switching events are the decisions made by the algorithms. A variation of the EDP is the power-delay product (PDP), which gives power consumed per switching event. PDP is a misleading metric compared to EDP because it favours a processor that operates at lower frequencies [212]. EDP gives the average energy multiplied by the time it takes to do the computation. Thus, EDP takes performance into account.

3.3.3.3 Area Efficiency

With the increase in lab-on-chip techniques, there is an ever increasing need to miniaturise platforms for clinical diagnostics. The need for miniaturisation is becoming inevitable because of the proliferation of chronically implanted devices at very invasive regions in the body. Miniaturisation ensures that the obstruction of normal internal and/or metabolic activity is mitigated [211]. Area efficiency is a metric that provides information on the useful area per operation in a microchip. It is obtained by dividing the number of operations by the chip area occupied to do those operations.

3.4 To Standardise or to Customise Performance Measures?

A major hurdle in the deployment of standalone BMI systems are the clinical trial costs involved in developing such systems. Due to their sensitive nature, medical devices require excessive regulatory oversight; well thought-through deployment strategy; substantial design and manufacturing cost and most importantly, a rigorous validation

process. The validation process starts prior to preclinical experiments, because for clinical translation to field, it is necessary to ensure that devices perform desired function with high certainty. A major bottleneck in validating algorithms are the choice of performance measures and metrics.

Accurate selection of appropriate performance metrics are the cornerstone for a balanced, objective and fair evaluation of detection algorithms. Performance evaluation is essential in BMI; however, it varies from application to application. Due to the multitude of non-standard performance measures in use for BMI, the performance of the same algorithm can be interpreted differently. For example, a single algorithm can have a 'high' measure in one metric and a 'low' measure in another metric, both of which intend to measure the same property. Inconsistencies like this further obfuscates the validation process. This is not far-fetched considering the rise in use of words like 'high/increased accuracy', 'low/reduced delay', 'low/reduced complexity' and so on to describe BMI performance. Words like these, instead of providing more insights, end up creating more unanswered questions like; what sort of accuracy or complexity metrics were adopted? How do the selected metrics compare to other adopted metrics? How beneficial are these measures/metrics to the application? How does this system compare to other current systems that use a different metric? All these questions could be avoided if standardised measures are used, such that every current work adopts a standard metric.

Another common issue that reinforces the argument for standardisation, is that performance measures with the same meaning sometimes have different names. A typical example of this is the classification accuracy sometimes called success rate, mean accuracy, non-error rate, probability of success and many more names. This inconsistency in nomenclature can lead to misinterpreting the metric and discrepancies in reporting performance. As against the argument for having standardised measures for BMI applications, customised measures can be created for various applications. Customised measures using weighted efficacy and complexity measures could be used to more accurately measure application specific efficiency or optimality. Since the efficacy needs of various applications are dissimilar as highlighted and issues like complexity affect applications differently depending on the need for chronic or non-chronic implantation; and how invasive, semi-invasive or non-invasive the implant is. For customising performance metrics, metrics that emphasise more on the current needs and trends of the industry can be selected. Currently, efficacy is of primary importance as most studies are proof-of-concept studies. Until BMIs achieve performance levels that allow patients accomplish routine activities unassisted, efficacy remains the most important requirement for BMI systems. Afterwards, focus can then be shifted to other metrics that asses complexity and efficiency. In spite of the breakthroughs in BMI, there is still a long way to go before they can find large scale application [213]. As stated earlier, their applicability is highly dependent on ease of use. Similarly, the adoption of different metrics for evaluating detection algorithms in BMI is dependent on their ease of measurement and how universal they are. In the rest of the work, performance measures are selected

based on the research issues addressed by each work. The unavailability of standard measures for BMI detection algorithms prompted the need for a thorough review and commentary on performance measures in BMI detection algorithms so that suitably tailored measures are selected that are fully representative and compatible with set performance targets.

3.5 Chapter Summary

As stated in Chapter 2, the major challenges of physiological signal analysis are: sensing devices, feedback signals and feedback algorithms. This was what prompted the focus on this three. For feedback algorithms, a major challenge is the unavailability of standard measures for evaluating feedback algorithms in closed-loop DBS and BMI applications. For feedback signals, signals are required that clearly distinguish disease and non-disease conditions and are consistent across time and patient. For sensing devices, characteristics like implantability, spatio-temporal resolution and invasiveness are essential. In the rest of the work, performance measures are selected based on guidance from this review. Below is a summary of what the chapter covered:

a) BMI are rehabilitation tools in which neural functions are modulated through feedback that is triggered by either decoded external percepts or brain activities. These are respectively representative of the two main categories of BMI. Category one controls perception to the brain based on external sensory information, while category two controls the actuation of a body part or a prosthesis based on recorded neural activity from the brain, of which closed-loop DBS is an example. The review focusses on category two BMIs, which are central to closed-loop DBS.

b) To facilitate closed-loop DBS, there are a wide range of neural signals that could be obtained using various signal acquisition techniques. Neural activity measurements consisting of metabolic and electrical activity are the preferred choice for use as feedback signals.

c) Metabolic activity measurements are more selective, specific and quantifiable than electrical activity measurements. On the other hand, electrical activity has a faster response than metabolic activity. Of all electrical activity measurements, LFP has the optimal trade-off in spatio-temporal resolution as well as stability. This makes it a prime candidate for closed-loop DBS.

d) For optimal performance, detection algorithms in BMI systems are required to have high efficacy, low complexity and high efficiency. To ascertain this requirements, the first challenge is to choose which metrics or measures to use as performance indicators. So far in BMI applications, performance requirements are to an extent dependent on the type of application. Because a measure that is evidently 'high' in one application may be evidently 'low' in some other application. Thus, standardising and customising metrics for evaluating performance remains a big challenge. Nevertheless, various applications have gravitated towards certain metrics and standards based on their simplicity and popularity of use in state of the art designs. An

example of such is the use of classification accuracy. For all its inconsistencies, the classification accuracy seems to be the most widely used metric for measuring efficacy. This can be attributed to its simple and straightforward nature – even though it is sometimes misleading. However, for a more accurate assessment of efficacy, more balanced measures like F1-score, MCC and many more customised metrics that incorporate one or more reliable efficacy measures could be used.

e) Finally, for BMIs to facilitate unrestricted interaction by patients, the goal remains real-time detection algorithms at the highest achievable accuracy (approaching 100%), and hardware efficient algorithms. However, current research is still miles away from this mark. The major bottleneck is the great scientific challenge of understanding how the brain works. Notwithstanding, there seems to be so much potential in BMI applications. Who knows, BMI applications could take a dramatic turn in fortunes such that the desired levels of performance are achieved in our lifetime.

Chapter 4

Towards On-Demand Deep Brain Stimulation Using Online Parkinson's Disease Prediction Driven by Dynamic Detection

4.1 Introduction

Clinical DBS for PD uses continuous, high frequency voltage or current pulses to mitigate PD. The major setbacks of present clinical DBS are stimulation induced side effects and shortening of pacemaker battery life [24], which can be addressed using on-demand DBS. It regulates stimulation by controlling the stimulation intensity and timing using feedback signals from the stimulation site [163]. On-demand DBS has the potential of improving stimulation efficacy, reducing power consumption and reducing side effects [24]. Several studies have proposed different ways of obtaining minimally invasive feedback signals for on-demand DBS. Primary focus has been placed on internal, e.g. LFP, ECoG [24, 163]; and external (e.g. EMG) electrophysiological signals [89]; biochemical signals [214]; and mechanical signals [89]; some of which have shown promising results. Electrophysiological signals obtained from external measurements are less suitable for detecting spontaneous PD activity due to their low correlation with rigidity and bradykinesia, [38, 215]. The discomfort experienced by patients due to externally attached sensors may also be an issue. Using biochemical signals, detection may not be instantaneous and may be hindered by the need for bulky and complex devices in some cases. Internal electrophysiological signals such as LFP are known to be indicative of PD symptoms such as rigidity, slowness of movement or tremor [16]. Also, recordings can be obtained from the same electrodes that are used for stimulation [115], making them minimally invasive. However, LFP correlations to bradykinesia and rigidity are functionally different from those of tremor [216]. This reinforces the need to identify distinguishing features in the data for individual patients. Various studies have mainly focused on monitoring beta band LFPs only [24, 82, 120]. However, using only beta band LFPs may not be sufficient, as they have not displayed satisfactory consistency across time and patients [24, 121]. Furthermore, the correlation of gamma [16, 122, 123],

and tremor [124] bands with PD symptoms, raises more questions on the suitability of using beta band activities alone. In essence, triggering DBS using characteristics from only a single band may be ineffective. Alternatively, to create robust (effective) feedback algorithms, identifying the most relevant recording channels (in multichannel recording) and frequency bands can provide a better mapping between LFP recordings and disease states. Hence, for every patient and at certain intervals of time, the channels and/or frequency bands that display the most pronounced variation between PD and non-PD events can be determined and used to detect PD states; this implies using fewer features, which are dynamically updated. The mapping between the LFP features and PD states may not be straightforward, which can cause the selected features to perform poorly. Detection accuracy can be improved using an additional dynamic stage consisting of dynamic classifiers. The processing chain is depicted in Figure 4.1. The implementation of a robust PD detection scheme is necessary because inaccurate detection results in administering stimulation when it is not required, and this may lead to stimulation induced side effects [217]. Or, inaccurate detection may result in the non-administering of stimulation when it is required, which may worsen patient condition [37].

This chapter examines and evaluates a subset of BMI algorithms suitable for on-chip implementation of PD detection in real time with high performance and low complexity. Computationally efficient on-chip (online) PD detection would facilitate the development of fully implantable closed-loop DBS systems that could automatically adjust stimulation parameters by the brain response in real time. The optimum combination of algorithms in terms of detection accuracy and computational complexity consisting of feature extraction (FE), DR and dynamic classifier algorithms are identified. A novel DR technique, the MRM is proposed. The algorithms are tested using representative PD and non-PD datasets to choose the best algorithm combination for real-time hardware implementation.

The chapter is organized as follows. Section 4.2 details the semi-synthetic LFP generation method. Section 4.3 describes the methodology for evaluating the algorithms. The

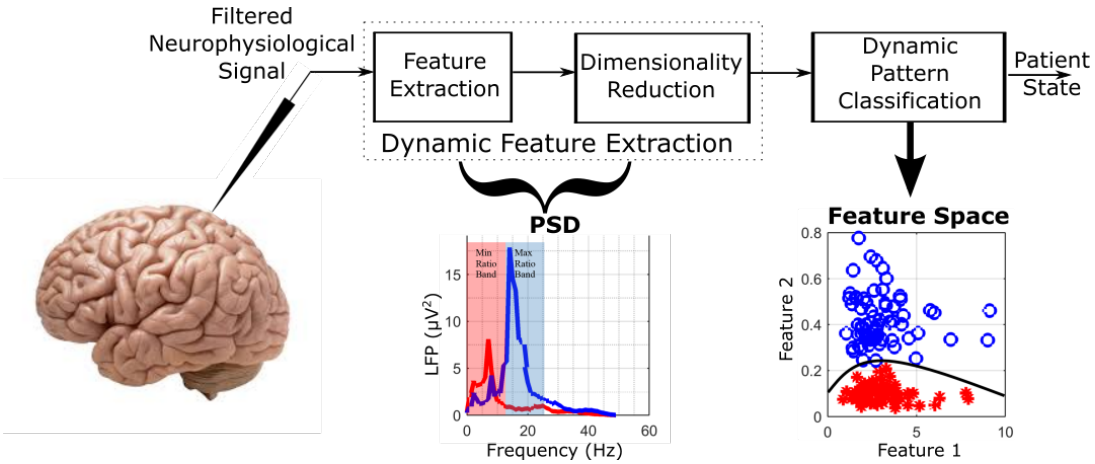


Figure 4.1: Typical bio-signal processing chain of PD state determination during on-demand DBS. PSD represents Power Spectral Density.

candidate algorithms for evaluation are described in Section 4.4. Section 4.5 compares the performance of the examined algorithms in terms of accuracy and complexity. Discussion and chapter summary are presented in Section 4.6 and Section 4.7 respectively.

4.2 LFP Generation

To evaluate the performance of the algorithms, accurately labelled LFP recordings for PD patients are needed. Studies have used modelled STN LFP recordings for controlling DBS [69, 82]. These may be inadequate since LFP recordings have been reported to be modulated by cognitive, emotional and behavioural tasks [16, 123, 218], which introduce unpredictable variations. Using real physiological recordings produces better models that incorporate the dynamic variations present in LFP signals. So far, studies in PD detection have been hindered by the unavailability of standard databases of Parkinsonian LFP signals that could be used for evaluating algorithms. A possible solution is to use LFP recordings from PD patients subjected to L-dopa, a common pharmacological therapy for PD patients, consisting of periods in which patients are ON and OFF L-dopa. The ON L-dopa periods are periods when L-dopa medication is effective and is normally accompanied by little or no PD symptoms, while OFF L-dopa periods are when PD symptoms return, signifying periods of motor deficit [219, 220]. The test datasets used were LFP recordings from the STN of subjects exhibiting a combination of bradykinesia and/or rigidity during the onset of PD, with less noticeable tremor. Recordings were made from nine patients with PD who had bilaterally implanted DBS electrodes in their STN and are referred to as dataset A – I. Each patient recording contained separate ON and OFF L-dopa data between 5 to 10 minutes long. The data was obtained from the Department of Clinical Neurology, University of Oxford. Recordings were made prior to the connection of a subcutaneous DBS pacemaker and stimulation was completely off during recording. Details on the daily drug dosage, on and off UPDRS score and dominant symptoms for eight of the nine patients are summarized in [24]. The permanent quadri-polar macro-electrode used was model 3389 (Medtronic Neurologic Division, Minneapolis, MN) consisting of 4 platinum-iridium cylindrical contacts. Its contacts are numbered 0, 1, 2 and 3, with 0 being the most caudal and 3 being the most cranial for both right and left electrodes – making a total of eight monopolar channels for each patient. ON and OFF L-dopa LFP data are referred to as non-PD and PD data respectively. Figure 4.2 shows a snapshot of OFF and ON L-dopa recordings of the left DBS lead for dataset A. In order to increase the length of recordings which mimic the unpredictable nature of LFP recordings, semi-synthetic datasets can be modelled using the LFP recordings from each dataset. There are a number of approaches to achieving this as summarised below.

4.2.1 Semi-Synthetic Data Generation Techniques

Semi-synthetic data generation provides the flexibility to manipulate the signal characteristics such that all underlying conditions are represented. This enables conclusions

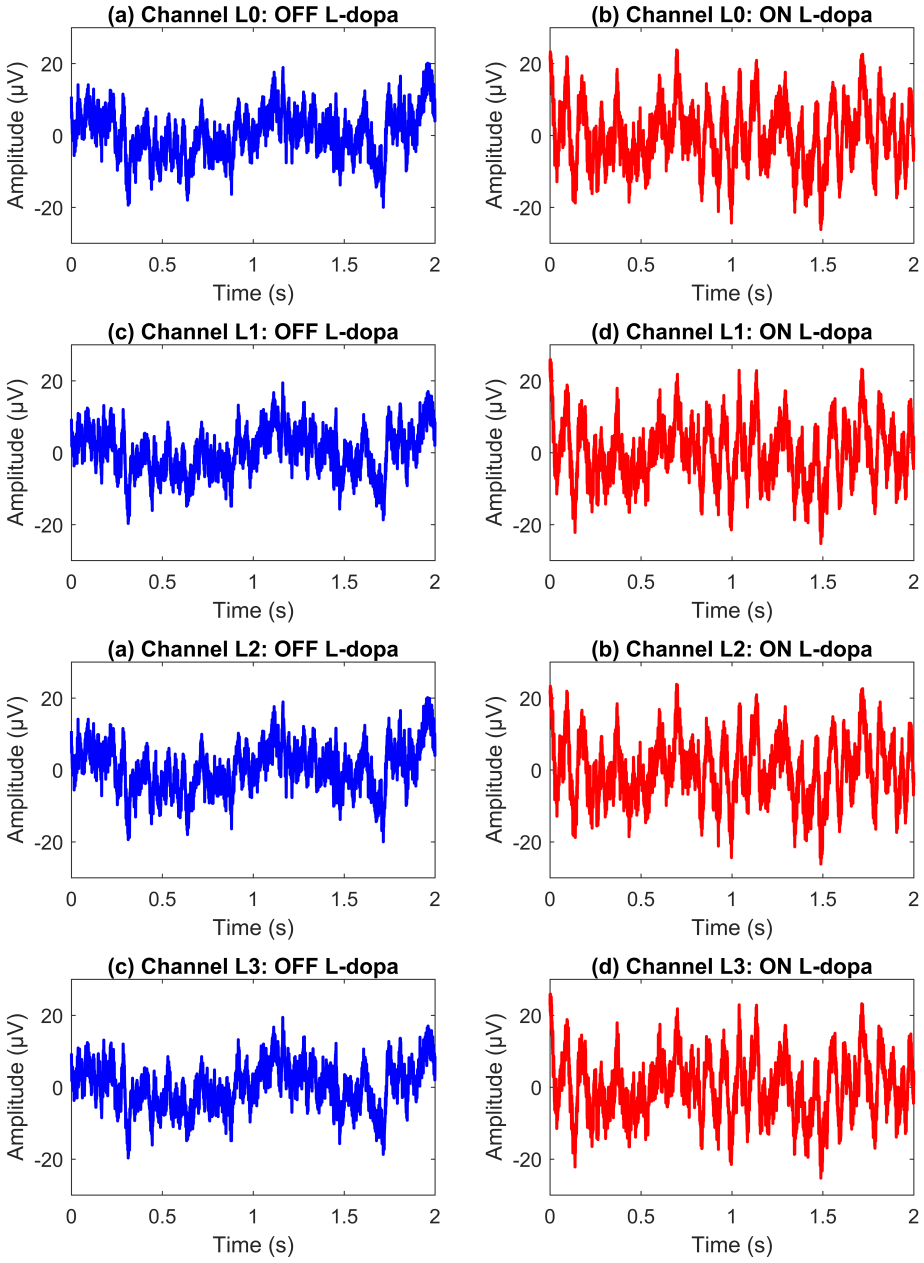


Figure 4.2: A snapshot of OFF and ON L-dopa recordings of the left DBS lead of dataset A.

that could be extended to an entire population. LFPs are extracellular activities obtained from a localised population of neurons, making it necessary to utilise inherent statistical properties that maintain the information modelled by individual samples, as well as segments of the signal. Nevertheless, there are many ways this could be done, depending on the property to be exploited. Some statistical properties could be exploited including, Gaussianity and stationarity of the signals. Below are a list of ways semi-synthetic data can be generated by exploiting statistical properties.

4.2.1.1 Resampling Based Methods

To use resampling methods for semi-synthetic data generation, the original LFP data is divided into epochs. The epochs are recycled randomly to create new LFP data which is representative of the underlying conditions. In statistics, resampling methods are used for many reasons, notable is their use for validating models. For use in validating models, the major techniques include random permutation sampling and data recycling methods like bootstrapping and jackknifing.

1. *Random permutation sampling*: In random permutation, LFP epochs are randomly ordered and are selected without repetition. Its major shortcomings are that the set to be selected from is required to be long enough since there are no repetitions. Also, the quality of its resampling is dependent on the major source of randomness, that is, its pseudorandom epoch selector.

2. *Data recycling*: Data recycling involves random sampling of epochs with replacement. The major methods are bootstrapping and jackknifing. In bootstrapping, data is selected from the same data pool randomly. In contrast to this, the jackknifing method deletes a selected number of observations say, N each time data is selected. A major advantage of data recycling is its simplicity in recreating new datasets. However, an apparent disadvantage is the possibility of data leakage, such that the training set leaks into the test data set.

Using resampling-based methods may not be suitable as the signals are not long enough to be used with random permutation sampling. For data recycling, a major issue is data leakage, which will provide results that may not be statistically valid.

4.2.1.2 Blind Source Separation (BSS)

For brain signals, the pattern of activity from statistically independent sources that contribute to the parent signal can be obtained using spatial source separation techniques. Contributions from various spatial sources can be modelled to create semi-synthetic data sources. This is pertinent because the brain models information in a statistical way using multi-physiological activity [221]. This consists of linear mixtures of signals from various statistically independent sources. Blind source separation (BSS) can be used to recover the sources of a signal. A major BSS technique is the independent component analysis (ICA). ICA is a method for finding the underlying components in statistical data. ICA does this by finding components that are statistically independent and non-gaussian. In most signals, the dominant form of non-gaussianity is sparsity. Sparsity in signals can manifest in time, frequency or space. Mathematically, ICA can be represented as follows [222],

$$x_i = \sum_{j=1}^{m} a_{ij} s_j \tag{4.1}$$

where s_j are the m-independent source signals $(s_1, s_2, .s_m)$ and x_i is a discrete point in the original signal, and a_{ij} is the mixing matrix. ICA tries to find approximations

for the mixing matrix based on assumptions that the various sources are statistically independent and non-Gaussian. ICA is a classic case of unsupervised learning. A major shortcoming of ICA is that it gives a number of components and does not provide sufficient information on the statistical significance of the components (in the case of neural signal separation, neural sources). Hyvarinen [222] proposed a method by which to overcome this shortcoming by using inter-subject consistency to assess the significance of a component. For the original signal to be broken into its component sources, it has to be non-gaussian and its sources s_j have to be statistically independent. For our original LFP signals, there is a need to investigate if they can be broken down into their independent components. This is necessary because LFP activities are activities from a localised population of neurons, and as earlier stated contributions from various spatial sources can be modelled to create semi-synthetic data sources. The first step is the visual inspection for non-gaussianity, which is summarised in Figure 4.3. From Figure 4.3, after detrending and filtering to within 50 Hz, the resulting signals are Gaussian, as such their independent components cannot be obtained. Thus we can investigate the use of stationarity to model the underlying conditions.

4.2.1.3 Autoregressive and Moving Average Models

In time series analysis, when successive observations are dependent, future values and underlying conditions could be predicted from past observations. In neural signals, due to the various interconnections between localised populations of neurons, there is a tendency for future neural activity to have some dependency on past observations (activity). All the hypothesis that model the brain dynamics have alluded to the fact that it has a non-linear dynamics [50–52]. And like all non-linear systems, the individual dynamics of its neurons varies from the superposition of the dynamics of the individual neurons. This makes time series data from the brain stochastic; the more exact models are called deterministic models. Stochastic processes are statistical phenomenon that evolves in time according to probabilistic laws [223]. The time series data may appear to be 'random', however, some of this randomness could be explained in terms of probability models, like the autoregressive (AR) or moving average (MA) models. This could be done for signals exhibiting stationarity. Stationarity could be strict sense or weak stationary. For AR, MA and autoregressive moving average (ARMA) processes the mathematical models that describe these are explained as follows.

Moving Average (MA) Process

In a stochastic process, a moving average process models current observations based on linear dependency on past values. As an example, a process X_t can be represented as a moving average process with order q [223],

$$X_t = \mu + \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \tag{4.2}$$

where μ is the mean of the series and $Z_{t-q}, Z_{t-q-1}, \ldots, Z_t$ are independent white noise error terms. Also, $\beta_0, \beta_1, \ldots, \beta_q$ are the parameters of the model.

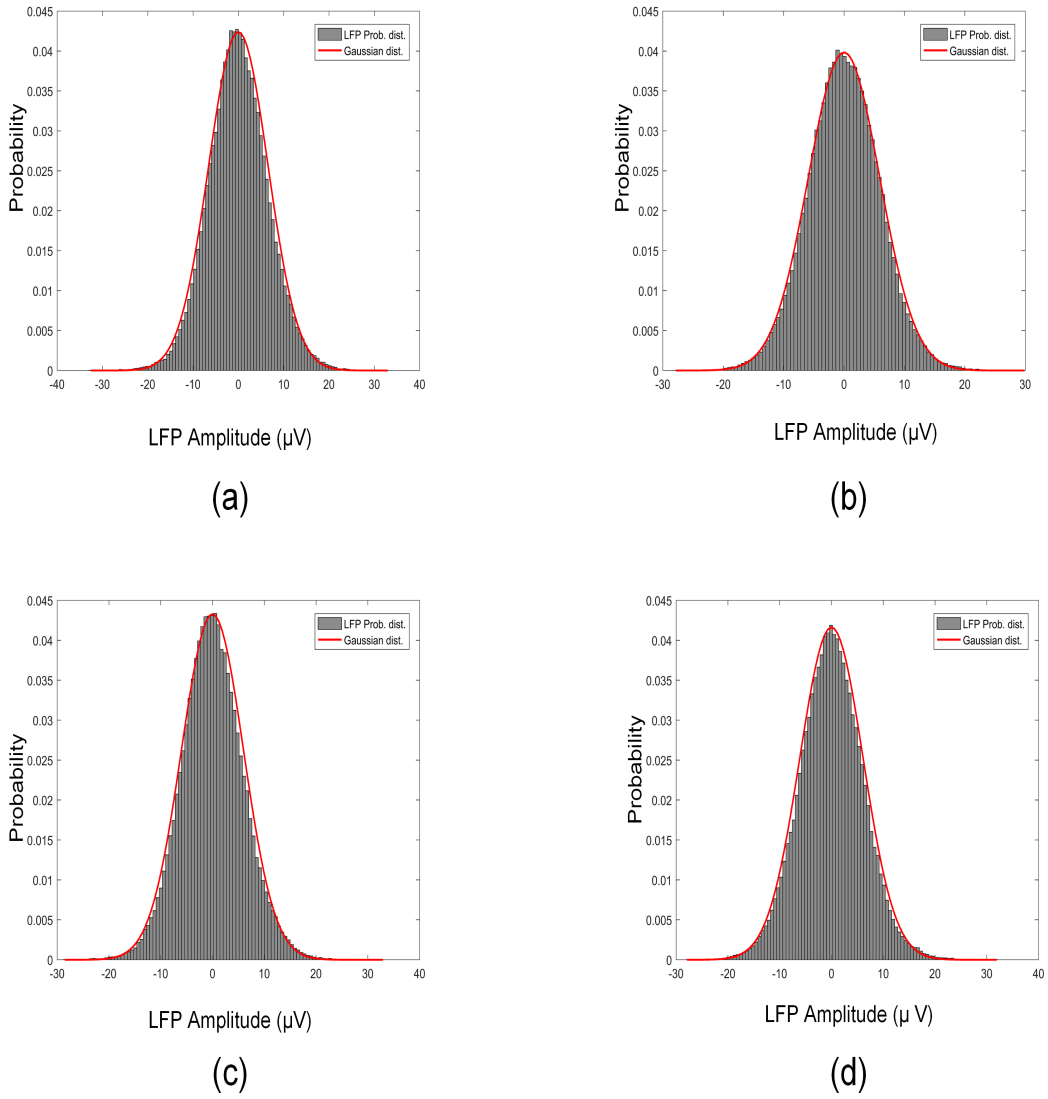


Figure 4.3: Probability distribution of LFP Amplitude for selected channel of dataset A. a) Channel L0. b) Channel R1. c) Channel L2. d) Channel R3.

Autoregressive (AR) Process

Another representation of stochastic process can be obtained using the AR process. Like the MA process, current observations are dependent on past observations, however in a slightly different way. An AR process X_t with order p can be represented as [223],

$$X_t = c + \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t \tag{4.3}$$

where c is a constant term, Z_t represents white noise; $\alpha_1, \ldots, \alpha_p$ are the parameters of the model. To use for forecasting, it has to be wide sense stationary (WSS). As an example, an AR(1) process is stationary if $\mid \alpha_1 \mid < 1$. In a broader sense, an AR (p) model is WSS if the roots of its polynomial [223],

$$z^p - \sum_{i=1}^{p} \alpha_i z^p = 0 \tag{4.4}$$

lie within a unit circle, that is, all its $\mid z_i \mid < 1$. The stationarity test will be discussed in later sections.

ARMA Process

Since AR processes can be used to represent WSS processes, ARMA models represent parsimonious alternatives of this representations. An ARMA (p, q) model has AR process of order p, and a MA process of order q [223]. As an example, an ARMA $(1, 1)$ can be represented as [223],

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + Z_t + \beta_1 Z_{t-1} \tag{4.5}$$

where the observations are $X_1, X_2, \ldots, X_K$, with X_t as the current observation and X_{t-1} as the previous observation. With μ as the mean, α_1 and β_1 are the AR and MA parameters respectively. While Z_t is a purely random process with mean zero and variance σ_Z^2.

Stationarity Test

In a broad sense, a time series is stationary if there is no systematic change in mean, variance, and if there are little or no periodic variations [223]. In essence, the statistical properties of the various sections of the data are more or less the same. But in the strict sense, it may be difficult for a time series data to be stationary. Stationarity can only be associated with a model, as such a stationary time series infers that the time series data exhibits properties that suggest that a stationary model could be 'reasonably' fitted to the data. A time series is strictly stationary if the joint distribution of $X(t_1), \ldots, X(t_k)$ is the same as the joint distribution of $X(t_1 + \tau), \ldots, X(t_k + \tau)$ for all $t_1, \ldots, t_k, \tau$. In essence, shifting the time series by a translation τ does not affect the joint distribution. However, for our signal, it is clear that the distribution $X(t)$ cannot be the same for all t. As such, we can explore the possibility of it being stationary in a less constrained way. For WSS, the following condition has to be satisfied [223],

$$E[X(t)] = \mu$$

$$Cov[X(t), X(t+\tau)] = \gamma(\tau) \tag{4.6}$$

where $E[X(t)]$ is the expectation (first moment) of the joint distribution $X(t)$. $Cov[X(t), X(t+\tau)]$ is the covariance (second moment) of $X(t)$ and $X(t+\tau)$. $X(t+\tau)$ is a time differenced version of $X(t)$. The Dickey-Fuller test can be used to test for non-stationarity in an autoregressive model, by testing wheather a unit root is present. However, since our signals are brain signals that have the possibility of having higher order AR(p) processes, then we will use the Augmented Dickey-Fuller (ADF) test for

unit root will be appropriate. The presence of a unit root signals non-stationarity, which could lead to difficulties in statistical inferencing. The ADF test assesses the null hypothesis of a unit root using the model [223],

$$y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \varepsilon_t \tag{4.7}$$

where Δ is the differencing operator, that is $\Delta y_t = y_t - y_{t-1}$, p is the number of lagged difference terms and ε_t is a mean zero innovation process. For the null hypothesis of a unit root, $\phi = 1$, and under the alternative hypothesis $\phi < 1$. If a test fails to reject the null hypothesis, then it fails to reject the possibility of a unit root. For non-stationary signals, without seasonality, an autoregressive integrated moving average model (ARIMA) can be used. While for seasonal non-stationary data, seasonal ARIMA models can be used. The original LFP recordings rejected the null hypothesis of a unit root against the autoregressive alternative. This makes them suitable for fitting AR processes. In place of AR processes, ARMA processes will be fitted as a stationary process may be adequately modelled using ARMA models consisting of fewer parameters than a pure AR or MA process on its own [223]. Thus, ARMA models would be used to synthesis new LFP recordings. ARMA models have been used in bio-electrical signals in [224].

4.2.2 LFP Data Synthesis

Table 4.1 summarises the various benefits and shortcomings of the highlighted data generation techniques. Based on Table 4.1, ARMA models were used to generate semi-synthetic datasets. Nevertheless, techniques from resampling methods were also incorporate in the LFP synthesis process. The following sections describe the complete procedure.

Table 4.1: Benefits and shortcomings of data generation techniques.

Approach	Benefits	Shortcomings
Resampling Methods	Simple and straightforward.	If datasets are not long enough, there is a tendency for data leakage between training and test datasets.
BSS	Modelling contribution from statistically independent sources is essential because the brain models information in a statistically sophisticated way using multilevel physiological activity [221].	Original data has to be non-Gaussian for ease in separation to its independent components. However, most neural signals are Gaussian [225], making it difficult to separate them into their independent sources.
MA, AR and ARMA	Takes advantage of probabilistic models in estimating the properties of a population from a sample.	Signals are required to be stationary or WSS to be used. Nevertheless, for non-stationary signal, a transformation could be used stationarise the data.

4.2.2.1 Estimating Parameters of the ARMA Models

For model fitting, the original LFP signal was low-pass filtered (10th order Butterworth filter, with cut-off frequency at 50Hz which is in the range of PD relevant LFP [123]), down-sampled to $fs = 128$ Hz (from 2048 Hz) and detrended. When dealing with ARMA models, the first step is to identify the order of AR(p) and MA(q) processes required, and then estimate the parameters of both processes. In determining the order of the AR and MA processes, visual inspection of the autocorrelation function (ACF) and the partial ACF (PACF) are useful. The PACF gives the partial correlation of a time series with its own lagged values by adjusting for the values of time series at all shorter lags [226]. As an example, given a time series y_t, if $\alpha(k)$ is the partial autocorrelation at lag k, then $\gamma(k)$ can be defined as the autocorrelation between y_t and y_{t+k} with the linear dependence of y_t on y_{t+1} through y_{t+k-1} removed. That is, it is the autocorrelation between y_t and y_{t+k} that does not completely account for lags 1 to $k-1$. Mathematically [226],

$$\gamma(k) = \begin{Bmatrix} Cor(y_{t+1}, y_t), & for\ k = 1 \\ Cor(y_{t+k} - P_{t,k}(y_{t+k}), y_t - P_{t,k}(y_t)), & for\ \geq 2 \end{Bmatrix} \tag{4.8}$$

where $Cor(y_{t+k}, y_t)$ is the autocorrelation between y_t and y_{t+k}, $P_{t,k}(y_t)$ is used for adjusting for the values of time series at all shorter lags and it denotes the projection of y_t onto the space spanned by $y_{t+1}, \ldots, y_{t+k-1}$. The PACF plays an important role in defining the order of an AR, MA or ARMA models. The ACF of AR processes is normally a mixture of sinusoidal and damped exponential functions. While that of an MA(q) process cuts off at lag q. In addition, the PACF of an AR(p) process cuts-off at lag p. On the order hand, PACF of an MA process will generally die out slowly. The AR and MA have reverse properties with respect to their ACF and PACF. In estimating the parameters, an iterative procedure is used. The residual sum of squares on a suitable grid of AR and MA parameter values are used. The point at which the curve plateaus, is the point at which additional parameters may be redundant. This represents the values which give the minimum sum of squares. This can be visualised using a plot of the residual sum of squares for corresponding values of p and q. This is termed residual analysis. Mathematically [223],

$$residual = observation - fitted\,value \tag{4.9}$$

In time series modelling, the observation is the original time series, the fitted value is the one step ahead forecast and the residual is the forecast error. The residuals are generally random for a model with a good fit. For a good model, more than 95% of the residual autocorrelations should lie within the range $\frac{\pm 2}{\sqrt{K}}$, where K is the number of observations. Ideally, the residuals are supposed to have an autocorrelation like that of white noise, such that there are no significant values present in the correlogram. In using Figure 4.4 to identify the order of the ARMA (p, q), both the ACF and PACF decay slowly. This means there is both AR and MA signature. The ACF cuts-off at a lag of 5, which means it requires an MA with an order of 5 or less. The PACF slowly

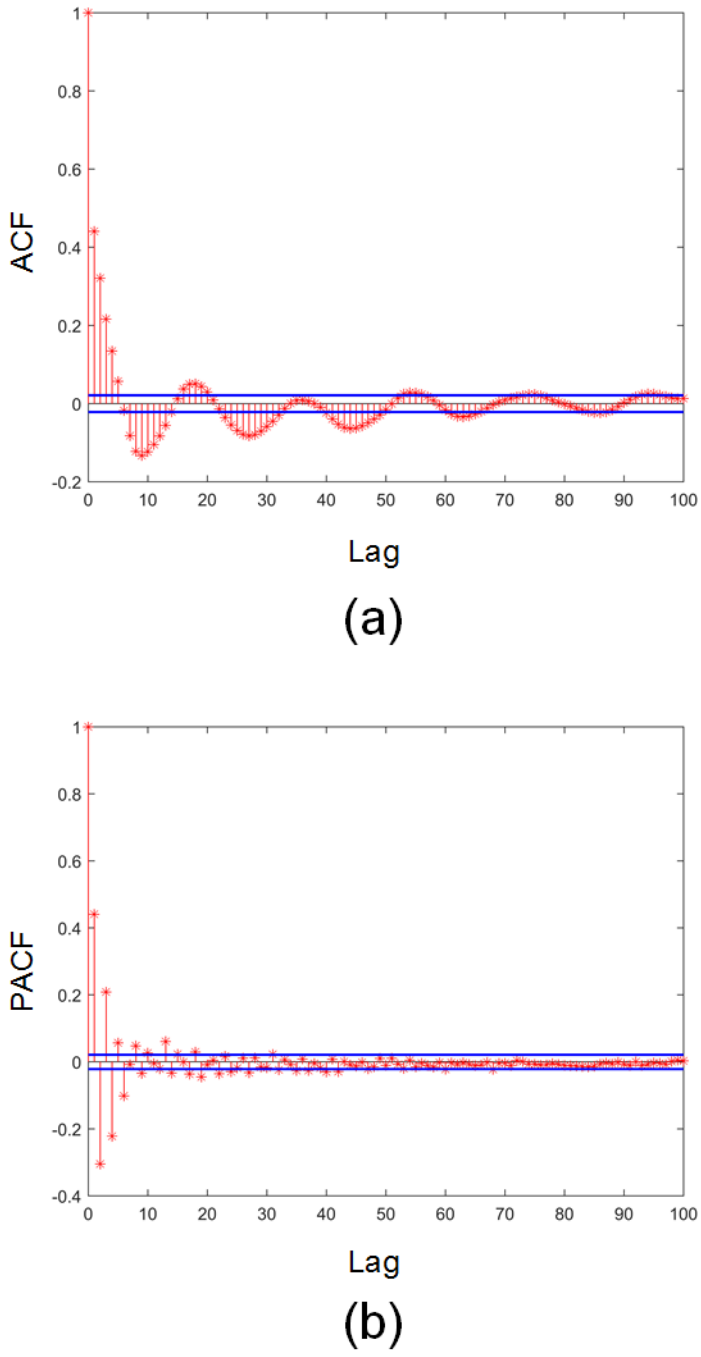
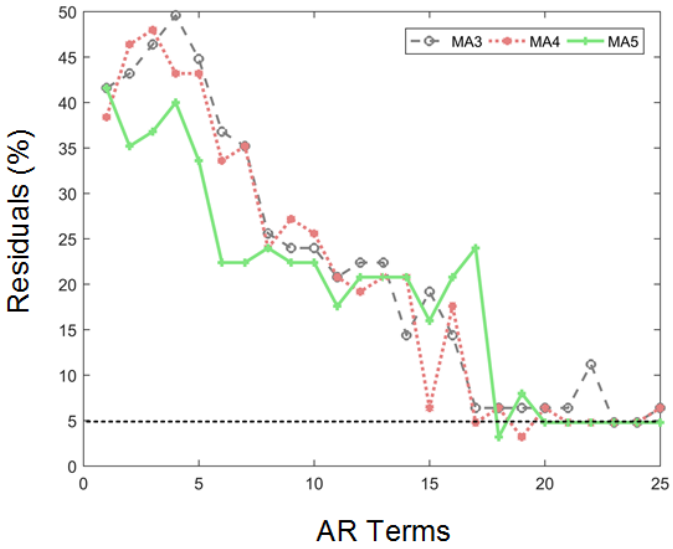


Figure 4.4: ACF and PACF for LFP signal of channel L0 in dataset A. a) ACF b) PACF.

decays, but it is not clear exactly when and where it cuts-off. So, a diagnostics test is run using BIC and residual analysis to ascertain the optimum order of the AR and MA processes.

The test determines the AR and MA pairs that give the lowest BIC. Also, the percentage of residual autocorrelations that are outside the error bound is essential; 5% residual autocorrelations is representative of residuals of white noise which is what is desired. As is shown in Figure 4.5, the combination 5% residual autocorrelations is obtained from ARMA $(18, 5)$ and ARMA $(19, 4)$. The model with the lowest BIC is chosen as it



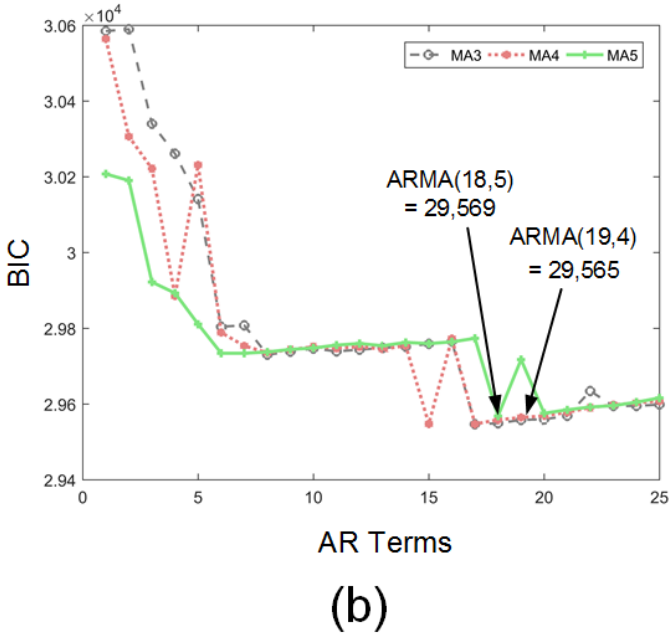


Figure 4.5: a) Residual for ARMA(p, q) terms. b) BIC for ARMA(p, q) terms.

represents the most parsimonious model. The differences between the two models are not fundamental which means both models can be used to fit the dataset.

4.2.2.2 Epoch Segmentation and Concatenation

To create semi-synthetic datasets with PD and non-PD episodes that imitate the progression of LFP signals in PD, segments of newly created LFP data need to be attached together to create long recordings. Segmentation is used to fragment continuous recordings into a series of epochs with a defined length. In addition to that in this context, they are used to subdivide the epochs into three regions based on their similarity. This segmentation aimed to ensure that any three succeeding PD epochs attached to create

long PD periods were selected from different segments, which introduces a distinctly different characteristic. Due to the difference in amplitudes between segments of the original LFP signal, similarity is measured using the normalized cross-correlation between epochs, which can be represented mathematically as [202]. ,

$$Corrnorm_{x,y} = \frac{\sum_{n=0}^{N-1} a[n]b[n]}{\sqrt{\sum_{n=0}^{N-1} a^2[n] \sum_{n=0}^{N-1} b^2[n]}} \tag{4.10}$$

where n is an integer, $0 \leq n \leq N-1$, and $a[n]$ and $b[n]$ are the two data sequences. Each epoch consists of 1 second PD or non-PD LFP data. As such, a 300-second LFP data consists of 300 epochs. So, for the data segmentation, a 300 epoch dataset is divided into three segments consisting of approximately 100 epochs (templates) per segment. All epochs in a segment are ensured to be nearly similar such that when synthesising long periods of PD or non-PD data, successively attached epochs are selected from different segments. This is important because it enables the assessment of the PD detection algorithms' robustness to instantaneous changes within the same patient state. In order to maintain all frequency components present in the original signal and to avoid introducing unwanted frequencies, successively attached epochs were slightly overlapped and averaged at the overlapping points [227]. To find the optimum overlap for successively attached epochs a similarity test was conducted. The test involved the original time-stamped LFP signal and newly created signals (not synthetic or semi-synthetic). The newly created signals were of the same length as the original signal and they were created by randomly attaching epochs to different section while using a specific number of overlapping samples. After which the normalised cross-correlation at lag zero of the frequency spectrum of the original signal and that of the newly created signal were found. Since the maximum value for normalised cross-correlation is one, the obtained value was subtracted from one to obtain the normalised cross-correlation error which represents the level of difference between the two signals (original and newly created). A sweep was conducted for samples covering between 0 to about 20% overlap. This was done for each channel of each dataset and the average of each dataset is found as tabulated in Table 4.2. Figure 4.6 shows a plot of the sweep for Dataset G. It can be seen that the local minima is at 14.0625%, which represents 18 samples overlap for 128 samples epoch. An overlap of 18 samples represents nine samples overlap at the beginning and end of each epoch. This was the process carried out in computing the optimum overlap in Table 4.2. From Table 4.2, the median overlap is 10.94% overlap, which represents 14 samples for a 128 sample/epoch.

4.2.2.3 Poisson Model for PD and non-PD Periods

PD and non-PD events contain point processes whose occurrence can follow a Poisson distribution. To create long recordings consisting of alternating PD and non-PD episodes that imitate the progression of LFP signals in PD; a Poisson distribution was used to define the duration of the PD and non-PD episodes. This was done to observe

Table 4.2: Optimum overlap for each dataset

Dataset	A	B	C	D	E	F	G	H	I
Optimum overlap (%)	10.94	20.31	1.56	10.94	20.31	1.56	14.06	1.56	1.56

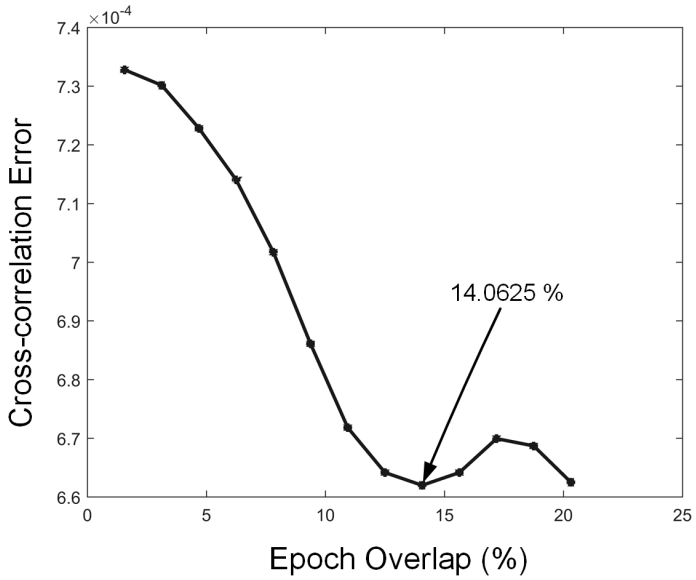


Figure 4.6: Normalised cross-correlation error for Dataset G.

how well the algorithm performed when tested with few and rampant PD episodes that mimic real-life situations [228, 229]. The duration of LFP signals was synthesised to imitate the non-linear progression common in PD which consists of three phases – improvement, stability and acute progressive decline [228]. The phases are defined as:

- PD improvement due to DBS intervention, with a mean PD rate (r) of 10% to 8%. That is, PD occurs 10% of the time (non-PD occurring 90% of the time) and it later reduces to 8% of the time. This represents a temporary improvement.
- Stability to PD rate (r) of 10%. From PD occurring 8% of the time it later deteriorates temporarily but stays within 10% for an extended period of time.
- Progressive decline as PD rate (r) increases in steps of 2% up 20% at its most advanced stage. After stabilising to 10%, it progressively deteriorates and occurrence increase gradually to 12%, 14% in steps of 2% up to about 20% at its most advanced stage. The 2% progression was chosen because motor disability score for L-dopa patients was reported to deteriorate annually by 2% [229].

The three highlighted stages were fitted within 12 hours duration, with each stage occurring about one-third of the time. The changes in PD rate (r), were defined over a 100 seconds (Δt) interval using the Poisson distribution below [230],

$$P\{d - PD\ events\ during\ \Delta t\} = \frac{e^{-r\Delta t}(r\Delta t)^n}{n!} \tag{4.11}$$

where, r is the mean PD rate, that is the proportion of time in which PD episodes are present. A PD event is defined to be one second, d is the number of PD events during Δt and where $r\Delta t$ is the average duration of PD events in an interval (Δt). Figure 4.7, shows example Poisson distributions for PD duration at each of the three phases of PD progression. Figure 4.7 (a) shows that of improvement in which proportion of PD duration reduces by two percentage points. Figure 4.7 (b) shows the stable PD stage in which PD events more or less occurs 10% of the time. Finally, Figure 4.7 (c) shows the third stage in which PD deteriorates rapidly until it reaches up to 20% average occurrence time at its most advanced stages.

4.2.2.4 Using ARMA Models for LFP Data Synthesis

The preceding sections describe the various techniques that are adopted for the LFP data synthesis process. Below is a brief description and summary of the complete process. For each channel in each dataset, a suitable ARMA (p, q) model was fitted to the original LFP signal – for both PD and non-PD data. An ARMA (p, q) model has autoregressive (AR) process of order p, and a moving-average (MA) process of order q [223]. For model fitting, the original LFP signal was low-pass filtered (10th order Butterworth filter, with cut-off frequency at 50Hz which is in the range of PD relevant LFP [123]), down-sampled to $fs = 128$ Hz (from 2048 Hz) and detrended. In selecting the appropriate order and parameters for the AR and MA terms, residual analysis was used. After fitting the appropriate models for both PD and non-PD data, longer recordings consisting of PD and non-PD periods are created. The process is summarized as follows:

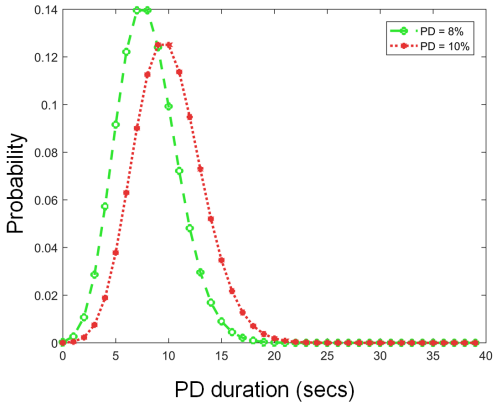
1. The PD data is firstly divided into one-second epochs. A 300 second PD recording thus consists of 300 epochs. These epochs are used as templates for the semi-synthetic data generation.

2. All the epochs are segmented into three regions based on their similarity – measured using the normalized cross-correlation between epochs. Thus, a 300 epoch dataset is divided into three segments consisting of approximately 100 epochs (templates) per segment. The aim of this segmentation was to ensure that any three succeeding PD epochs attached to create long PD periods were selected from different segments, which introduces a distinctly different characteristic.

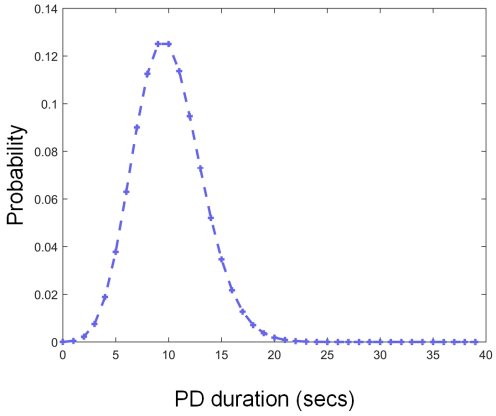
3. Using the fitted ARMA models, inferred residuals and individual epochs, a number of possible forecasts can be realised. Using an original epoch, Figure 4.8 shows forecasts from two possible Monte Carlo paths taken by a semi-synthetic epoch. For each one second original epoch, 100 Monte Carlo forecasts (with one-second duration) are made. Thus, for LFP recordings consisting of 300 templates, 30,000 semi-synthetic epochs were created.

4. Steps 1 – 3 are repeated for the non-PD data.

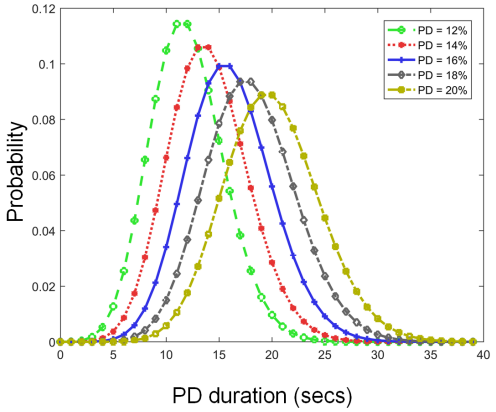
5. To create long recordings consisting of alternating PD and non-PD episodes that imitate the progression of LFP signals in PD; a Poisson distribution was used to



(a)



(b)



(c)

Figure 4.7: Poisson distribution for PD duration used for synthesing non-linear progression in PD. a) PD Improvement. b) PD Stability. c) PD deterioration.

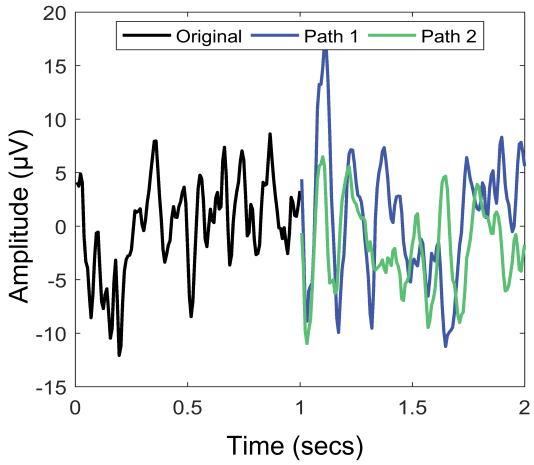


Figure 4.8: Semi-synthetic LFP generation showing an original epoch and two of its possible forecasted paths based on Monte Carlo simulations.

define the duration of the PD and non-PD episodes. This was done to observe how well the algorithm performed when tested with few and rampant PD episodes that mimic real-life situations [228,229]. The original and newly generated epochs in steps 1- 4 were used to synthesise 12-hour long LFP recordings.

During LFP synthesis, random permutation sampling was used for epoch selection. To maintain all frequency components present in the original signal and to avoid introducing unwanted frequencies, successively attached epochs were slightly overlapped and averaged at the overlapping points [227]. Figure 4.9 (a) shows a sample of the synthetically generated LFP recordings indicating PD (blue) and non-PD (red) periods. Figure 4.9 (b) shows the overlapping point averaging procedure. From Figure 4.9(b) it can be seen that the data overlap and averaging avoids the introduction of unwanted frequencies. The encircled regions of Figure 4.9 (b) depict this clearly: if concatenated at circle one without averaging, high-frequency components will be introduced while concatenating at circle two introduces low-frequency components. The overlapping and averaging of overlapping points serves as a trade-off to avoid introducing unwanted frequencies. Portions of the synthetic LFP were used to train the system to detect the patient state for unseen recordings. During FE, in instances where the sliding window selects segments of LFP data that have nearly equal proportions of non-PD and PD, are tagged as transition states.

4.3 Evaluation Metrics and Assumptions

4.3.1 Complexity Estimation

Complexity is quantified using a weighted cost of the number of operations (NOP) and the estimated microchip area. The smallest unit for the NOP is a 1-bit addition. Subtraction is considered to be equal to addition. Each multiplication or division is considered to be L-additions, using a quantisation of L bits/sample. Microchip area is divided into logic and memory area. Adders, subtractors and comparators are assumed

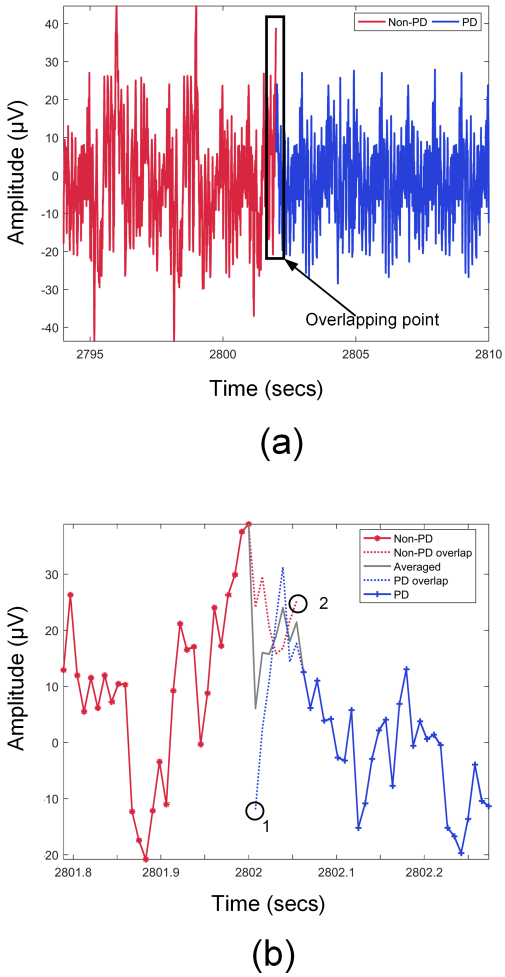


Figure 4.9: Semi-synthetic LFP generation showing a) Epoch concatenation. b) Averaging at overlapping point.

to be the same size. Following the procedure in [231] for a 90 nm CMOS process, each 1-bit adder is estimated to be $20.46\mu\ m^2$. The size of multipliers and dividers are assumed L-times the size of adders for L bits/sample quantisation. Memory is calculated based on the number of registers needed for each computation. A 1-bit register size was reported to be $15\mu\ m^2$ for a 90 nm CMOS process [231]. By labelling a combination of PD detection algorithm as a, its complexity cost is $CompCost_a$. The weighted complexity cost can be represented as follows,

$$CompCost_a = \frac{0.5 \cdot NOP_a}{max(NOP_1, NOP_2, ...NOP_v)} + \frac{0.5 \cdot area_a}{max(area_1, area_2, ...area_v)} \tag{4.12}$$

where NOP_a and $area_a$ are respectively the NOP and area for the combination labelled a. The number of combinations of algorithms is v and max(...) computes the maximum value of the various combinations. The complexity cost assigns a 50% weight to both NOP and area. A maximum complexity cost of 1 can be obtained for a combination concurrently having the largest area and the largest number of operations. Table

4.3 summarises the various costs used in evaluating the algorithms.

Table 4.3: Metrics for evaluating algorithms.

Error Cost	Complexity Cost	Combined Cost
F1-score	NOP	Error Cost
Choice Probability	Area	Complexity Cost

4.3.2 Detection Accuracy Calculations

The PD detection scheme consisting of feature extraction (FE), dimensionality reduction (DR) and dynamic pattern classification, were tested in MATLAB. Combinations consisting of all the three stages in the PD detection scheme were evaluated using the algorithms in Table 4.4. Each combination was evaluated using the following metrics which were discussed in Chapter 3.:

1. Classification error.
2. F1-score.
3. Choice probability (CP).
4. And a customised metric, the "Error cost". Each complexity cost ($CompCost_a$), has a corresponding error cost ($ErrCost_a$) that is calculated at 10% classification error and is

$$ErrCost_n = \frac{(1 - F_1) + (1 - C_P)}{2} \tag{4.13}$$

   A maximum error cost of 1 can be obtained for a combination having an F1-score of 0 and a CP of 0.

4.3.3 Assumptions

The algorithms evaluated were chosen based on their efficiency in previous BMI and bio-signal processing applications. The process involved evaluating a subset of FE, DR and machine learning models used in pattern classification. The detection scheme uses ON and OFF L-dopa signals as representative data for non-PD and PD cases respectively. It is also assumed that training occurs only once a day.

Table 4.4: The various algorithms to be evaluated.

Dynamic Feature Extraction		Dynamic Pattern
Feature Extraction (FE)	Dimensionality Reduction (DR)	Classifiers
DWT	MRM	Dynamic k-NN
STFT	PCA	Dynamic LR
	NMF	Dynamic SVM

4.4 Algorithms

4.4.1 Feature Extraction (FE)

In order to extract meaningful information, acquired physiological data which is normally in the time domain, is transformed to a computationally efficient form for further processing. In applications like PD state detection where power at certain frequencies can serve as biomarkers that indicate pathological states, time-frequency analysis is required. Frequency data provides information on where the power is concentrated for each pathological state, and the time domain data provides the instant they occur. Short time Fourier transform (STFT) and discrete wavelet transform (DWT) are examined and compared.

4.4.1.1 Short-Time Fourier Transform (STFT)

STFT uses the fast Fourier Transform (FFT) to obtain time-frequency data. This is achieved by dividing the signal into windows and FFT is applied to each window [232]. STFT is given by [232],

$$X_n[t;f] = \sum_{n=0}^{W-1} x[t+n]e^{-j\frac{2\pi}{W}nf} \tag{4.14}$$

where $x[t+n]$ is the discrete input signal at t which is the discrete time index, W is the window length into which the signal is split and f is the discrete frequency index. For this application, the time-stamped measurements are split into 2 seconds overlapping epochs, with 50% overlap between epochs. Also, the power bands (features) are divided into 5 Hz bands, with 3 Hz overlap between bands; 0 to 5 Hz, 3 to 8 Hz, ... 45 to 50 Hz. This provides a total of 16 features. The window is chosen such that a balance between time and frequency resolution is obtained.

4.4.1.2 Discrete Wavelet Transform (DWT)

It is a time-frequency representation that uses multi-resolution transformation. Mathematically DWT is given by [233],

$$W(u,2^j) = \sum_{n=-\infty}^{\infty} x(n)\cdot\Psi(\frac{n-u}{2^j}) \tag{4.15}$$

where $x(n)$ is the input signal, u is the translational parameter representing the time axis, 2^j (j is an integer) is the scale parameter representing the frequency axis and Ψ is the wavelet function [233]. Based on the scale parameter in Equation 4.15, at each level (j), it is down-sampled by 2 to the power of that level (2^j). For the DWT, a 4-level decomposition using the Haar wavelet was obtained. The average power at each level of decomposition consisting of 2 seconds overlapping epochs, with 50% overlap between epochs are obtained as features. The five features are defined as detail coefficient level 1 (32 - 49.5 Hz), level 2 (16 - 32 Hz), level 3 (8 - 16 Hz), level 4 (4 - 8 Hz) and approximation coefficient level 4 (0.5 - 4 Hz).

4.4.2 Dimensionality Reduction (DR)

To increase processing speed and computational efficiency, extracted features are compressed using additional processing termed dimensionality reduction (DR). DR involves reducing the number of features that will be used for patient state detection. A high number of features increases the possibility of data over-fitting, which results in a poor generalisation of unseen data. Also, periodically changing the extracted features used results in dynamic FE. This work explores linear DR methods, and proposes the use of maximum ratio method (MRM) (which is eventually chosen for this work).

4.4.2.1 Linear DR Methods

In BMI applications, the most primitive dimensionality reduction technique is feature selection. Feature selection is implemented by selectively retaining the most relevant features. Kung and colleagues [63], proposed feature selection in genomic signal processing whereby redundant genes were identified to sparsify the feature vectors. In another study on epilepsy seizure detection using EEG signals, Lee et al. [64], adopted feature selection by channel reduction. Channel reduction was implemented by manually identifying redundant channels, after which they were eliminated. Apart from manual DR techniques, to ease computation for on-chip implementation and because we intend to employ machine learning models for classification, we will be concerned with linear DR methods. Linear or projective DR methods can be represented by [85],

$$A = U^T \cdot B \tag{4.16}$$

where A is the reduced feature vector, B is the original feature vector and U is the projection matrix. All linear DR methods use the same set-up, nevertheless they differ on the criteria for choosing the projection matrix (U). The technique used in choosing the projection matrix determines the algorithms. However, due to the evolving nature of neural signals, linear DR methods may not be ideal for fully implantable applications. This is because there is a need to update the projection matrix periodically as the signal changes. Two linear DR methods are suitable for use with the proposed feature extraction methods: principal component analysis (PCA) and non-negative matrix factorisation (NMF).

Principal Component Analysis (PCA)

In PCA, the orthogonal basis (or principal components) that indicates the principal directions in which data varies is calculated [85]. In Equation 4.17, high dimension features X can be reduced to low dimension features A,

$$A = PC_i^T \cdot B \tag{4.17}$$

where $B \, \epsilon \, \mathbb{R}^{m \times n}$ consists of the training examples, m is the number of training features and n is the feature vector dimensions. Additionally, $PC \, \epsilon \, \mathbb{R}^{n \times n}$ contains the coefficients of all principal components, $PC_i \, \epsilon \, \mathbb{R}^{n \times i}$ contains the coefficients of the principal component up to the i-th principal component, making i the reduced feature vector

dimension and $Z \epsilon \mathbb{R}^{i \times m}$ consists of the new training features with reduced dimensions.

Non-negative Matrix Factorisation (NMF)

Since the features are extracted (power features) at the FE stage are all non-negative, NMF will be ideally suited for DR. It is similar to PCA in that it assigns weights to a set of bases to blend a representative observation. However, unlike PCA, all its weights are constrained to be positive. Where PCA gives negative weights, it gives zero weights, and this produces a sparse representation of the data assuming it has an underlying structure [234]. NMF can be represented by [234],

$$A \cdot H = B \tag{4.18}$$

where $A \epsilon \mathbb{R}^{m \times i}$ is the reduced feature vector, $B \epsilon \mathbb{R}^{m \times n}$ is the original feature vector. The matrix $A \epsilon \mathbb{R}^{m \times i}$ and $H \epsilon \mathbb{R}^{i \times n}$ are two smaller matrices which when multiplied approximately reconstruct $B \epsilon \mathbb{R}^{m \times n}$.

Preliminary Analysis of Linear DR Methods

To conduct a thorough analysis, preliminary analysis may need to be conducted such that one of the projective DR methods (PCA or NMF) is dropped. Using both algorithms increases the possible combination of detection algorithms consisting of FE, DR and classification. Since both methods project unto a subspace, their performance will be assessed based on their ability to retain the existing variations in the signals even after projection. The metric that will be used to assess this is the projection (or reconstruction) error, which was represented in Chapter 3.

The projection error for various features on each channel of each dataset was found. The average projection error was computed and is represented in Figure 4.10 (a) using DWT and Figure 4.10 (b) using STFT for feature extraction. From Figure 4.10 (a), it can be seen that the NMF has less reconstruction error compared to the PCA when one or two features are used, while the PCA does better when three or four features (zoomed portion) are used. The performance for the DWT does not provide conclusive evidence on the superiority of the two methods for use on the dataset. However, Figure 4.10 (b) presents a more consistent performance for the PCA compared to the NMF. Even though the NMF does better for features less than five, however, the PCA has a gradual increase in performance compared to the NMF. It is clear from Figure 4.10 (b) that the NMF has more variations in terms of projection error. This makes PCA a more trusted DR technique that is resistant to noisy data. Based on this preliminary analysis, the PCA will be adopted.

In addition to feature dimension reduction, the PCA will be used for channel selection since there are eight recording channels and only activity from a single channel may be necessary to enable classification. This is necessary because PD tracking using recordings from all channels may introduce redundancy as recordings from various channels might provide nearly similar information, as can be seen in the spectrum of the channels in Figure 4.11 which have nearly similar characteristics. On the other

hand, randomly selecting any one of the channels for FE can be counterproductive, as the channel with the least variation between PD and non-PD clusters may be selected.

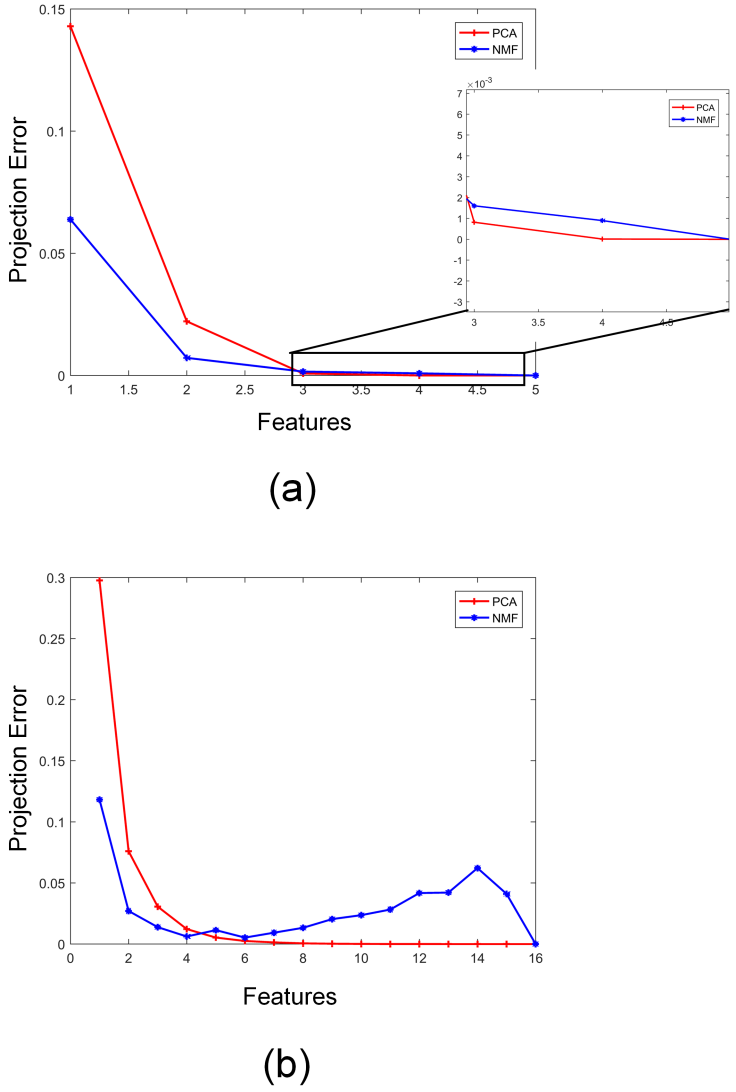


Figure 4.10: DR projection error using: a) DWT for FE. b) STFT for FE.

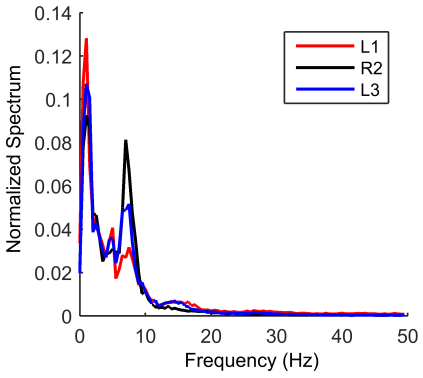


Figure 4.11: The power spectrum (for PD dataset) of some of the channels in dataset C, shown having nearly similar characteristics. L or R for left or right electrodes respectively, which are numbered from 0 to 3 (caudal to cranial contacts).

This necessitates the use of a methodical approach for channel and feature selection. For PCA, the channel with the least two-dimensional projection error is selected for use.

4.4.2.2 Maximum Ratio Method

Drawing upon the principles of Fisher separability criterion, the properties below are necessary to select features that produce the best separation in binary classification [85]:

- Features must be oriented such that they maximise the Euclidean distance between class means in the feature space. That is, maximising interclass distance is important for separability.
- Also, they must contain few outliers such that their within class variance in the feature space is minimised. That is, minimising intra-class distance is important for separability.

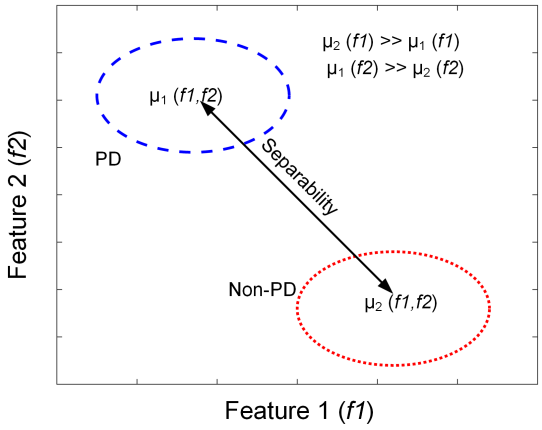
Unlike linear DR methods, the novel MRM is a DR method that uses labelled samples during training. It takes advantage of Fisher's criterion for separability. Since there is no need to quantify feature and channel separability in absolute terms, the MRM method finds a simple mathematical method for ranking the features and channels in terms of separability. The approach is targeted for hardware-aware implementations for DR. The process of MRM starts by identifying the channel having features with the most pronounced variation in activity. The goal is to obtain the feature space depicted in Figure 4.12 which makes classification easier. Figure 4.12 gives the most desired orientation for the clusters. The following section provides an analysis of the MRM and then details the implementation.

Analysis

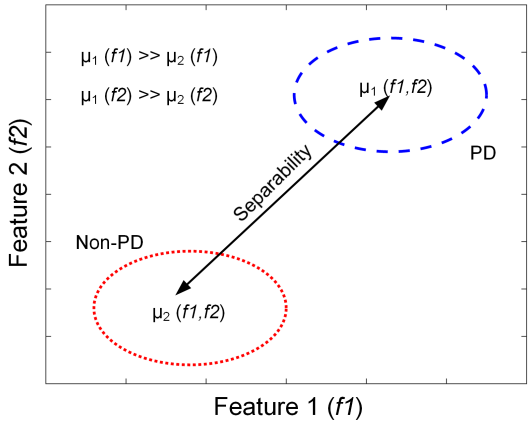
The above properties as highlighted by Fisher [85], are essential for classification because the between class distance represents how separable PD events are from non-PD events; while the within class variance represents how clustered PD and non-PD features are. These two properties are vital in making classification easier. Given a vector of training examples $O = [o_1, o_2, \ldots, o_D]^T$ having a total of D observations, with each observation (x_i) represented by n-dimensional features such that $o_i = [o_i^{(1)}, o_i^{(2)}, \ldots, o_i^{(n)}]$. This means O is a $D \times n$ matrix, where each observation o_i is an n-dimensional vector. If each of the observations belongs to one of classes in, $C = [C_0, C_1, \ldots, C_k]$ with each class having a total number observations $D_0, D_1, \ldots, D_k$ respectively (where total observations $D = D_0 + D_1 + D_2 \cdots + D_k$) . The mean μ_i and variance σ_i^2 for class C_i can be obtained as follows [85],

$$\mu_i = \frac{1}{D_i} \sum_{O \epsilon C_i} O \tag{4.19}$$

$$\sigma_i^2 = \frac{1}{D_i} \sum_{O \epsilon C_i} (O - \mu_i)^2 \tag{4.20}$$



(a)



(b)

Figure 4.12: Two-dimensional depiction of the desired orientation for the clusters (PD and non-PD) to create the largest variation using two features (f1, f2). (a) Scenario 1: feature space formed when one of the clusters (Non-PD in this case) has much higher f1 values, and the other cluster (PD in this case) has much higher f2 values. (b) Scenario 2: feature space formed when one of the clusters (PD in this case) has much higher values for both f1 and f2.

The Fisher's separability criterion, is inspired by the idea that separability is improved when the spatial distribution of observations from differing classes is high, while the spatial distribution of observations from the same class is low. That is, the criterion attempts to maximise the Euclidean distance between classes. To avoid features with a wider range to dominate those with a smaller range, such that each feature contributes approximately proportionately to its distance, the Euclidean distance is normalised by the variances for each feature [235]. For binary classification, we would like to maximise the ratio between the interclass distance (S_D) and intra-class variance (S_V) can be represented as follows [85],

$$\frac{S_D}{S_V} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{4.21}$$

where the mean for class i is $\mu_i = [\mu_i^{(1)}, \mu_i^{(2)}, \ldots, \mu_i^{(n)}]$ for n-dimensional features, with $\sigma_i = [\sigma_i^{(1)}, \sigma_i^{(2)}, \ldots, \sigma_i^{(n)}]$. Maximising separability translates to improved classifiability of the data. For binary classes the Fisher criterion $J(S_D, S_V)$ for the n-dimensional features can be represented as,

$$J(S_D, S_V) = max[\frac{(\mu_1^{(1)} - \mu_2^{(1)})^2}{(\sigma_1^{(1)})^2 + (\sigma_2^{(1)})^2}, \frac{(\mu_1^{(2)} - \mu_2^{(2)})^2}{(\sigma_1^{(2)})^2 + (\sigma_2^{(2)})^2}, \ldots, \frac{(\mu_1^{(n)} - \mu_2^{(n)})^2}{(\sigma_1^{(n)})^2 + (\sigma_2^{(n)})^2}] \tag{4.22}$$

where μ_1 and μ_2 are the mean values for the feature vectors in class one and two (since it is a binary classification problem). The respective variances for classes with mean μ_1 and μ_2 are given by σ_1^2 and σ_2^2. The features that maximise $J(S_D, S_V)$ represent the most suitable for classification. Assuming a 2-D feature space,

$$\sigma_i = [\sigma_i^{(1)}, \sigma_i^{(2)}]$$

$$\mu_i = [\mu_i^{(1)}, \mu_i^{(2)}] \tag{4.23}$$

Since the original LFP signals are Gaussian in nature, with mean values around zero in various frequency bands; and the power characteristics of LFP signals in different bands are used as features, the mean values are proportional to the standard deviations as represented by Equation 4.24 and depicted in Figure 4.13 which represents a 2-D feature space of dataset B obtained using MRM.

$$\mu_i \propto \sigma_i \tag{4.24}$$

where μ_i is the mean for class C_i, σ_i is standard deviation. From Figure 4.13, it can be seen that the means ($\mu_1$and μ_2) are approximately three times the standard deviations

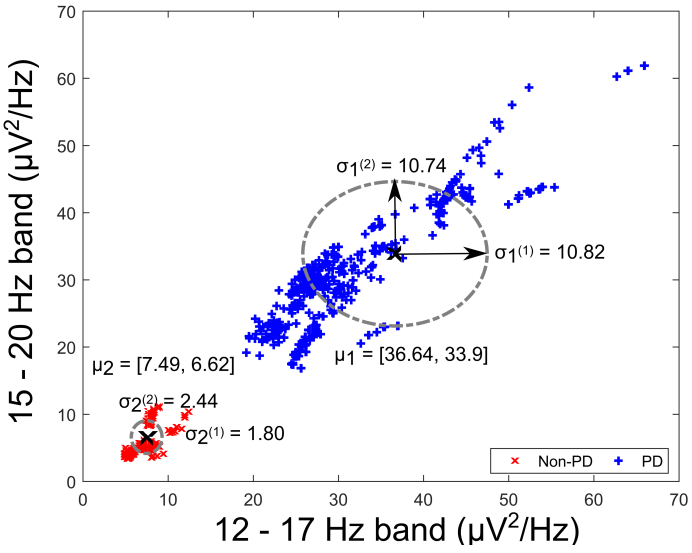


Figure 4.13: Feature space depicting mean and standard deviation of clusters for Dataset B.

(σ_1 and σ_2). For the objective function in Equation 4.22, rescales the Euclidean distance using the class variances such that feature separability is assessed relative to feature variance rather than the absolute Euclidean distance. In machine learning, there are different methods of feature scaling, it could be done using the or variances, as well as the mean value of the feature [235]. In our case in Figure 4.13 is can be seen that the standard deviation (and variance) for each feature is proportional to the means of the two classes. Instead of using the variance to rescale in Equation 4.22, the means of one of the two classes ($\mu_1^{(i)}$ and $\mu_2^{(i)}$) can be used to rescale the Euclidean norm in Equation 4.22. Assuming $\mu_1^{(i)}$ is the greater of the two, if $(\mu_2^{(i)})^2$ is used to rescale the euclidean norm, for the i-th feature, then Equation 4.22 is transformed to,

$$J(S_D, S_V) = max[(\frac{\mu_1^{(1)}}{\mu_2^{(1)}} - 1)^2, (\frac{\mu_1^{(2)}}{\mu_2^{(2)}} - 1), \ldots, (\frac{\mu_1^{(n)}}{\mu_2^{(n)}} - 1)^2] \tag{4.25}$$

where $\frac{\mu_1^{(i)}}{\mu_2^{(i)}} > 1$. This can be further represented as,

$$J(S_D, S_V) = max[\frac{\mu_1^{(1)}}{\mu_2^{(1)}}, \frac{\mu_1^{(2)}}{\mu_2^{(2)}}, \ldots, \frac{\mu_1^{(n)}}{\mu_2^{(n)}}] \tag{4.26}$$

Equation 4.26 represents the MRM which can be used to obtain an approximate of ranking the n-dimensional features from the most separable to the least separable. Ranking the features in terms of separability is necessary because using the most separable features for classification makes classification more accurate and less computationally intensive. The ranking can be represented more visually by Figure 4.14. In Figure 4.14, a sweep is conducted for the x-axis values in Figure 4.13, that is, $\mu_1(f1)$ ranging from $\mu_{1x} - \sigma_{1x} \rightarrow \mu_{1x} + \sigma_{1x}$ and for $\mu_2(f1)$ for values between $\mu_{2x} - \sigma_{2x} \rightarrow \mu_{2x} + \sigma_{2x}$.
Figure 4.13 above, then Figure 4.14 represents the plot of the cost function obtained using Fisher's separability criterion and MRM. It can be seen that the MRM and Fisher separability criterion have nearly similar patterns for the cost function for corresponding values of $\mu_1(f1)$ and $\mu_2(f2)$. The MRM approximates to the Fisher criterion in ranking features. Even though the exact values of the cost function are different. However, since features are ranked and arranged in relative terms, the absolute values in terms of the cost function may not necessary be the same, but it is important that the ranking are the same as shown by the surface plot in Figure 4.14. By taking advantage of the characteristic of our features (i.e., class means are proportional to class variances), the MRM estimates the feature ranks in terms of separability for the various features.
Aside from feature selection, the MRM can be used for channel selection. Based on the value of the separability cost function of the features for different channels, a weighted metric is used to assess separability for various channels. Channels are assessed based on a weighted separability cost function, with higher weights provided to the features with the highest ranks for each channel. The cumulative weights for each channel are used to assess separability by channel. The complete implementation for the MRM is summarised in the next section.

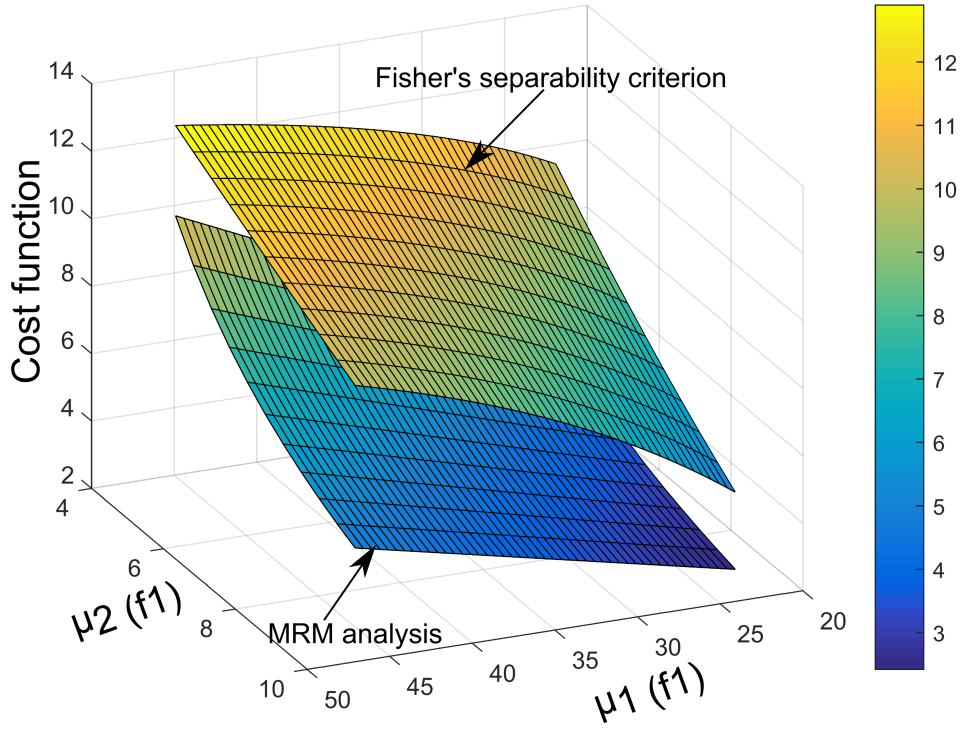


Figure 4.14: Cost function obtained using Fisher criterion and MRM for feature (f1) in Figure 4.13.

Implementation

Unlike PCA, the novel MRM is a DR method that uses labelled samples during training. The process of MRM starts by identifying the channel having features with the most pronounced variation in activity. The goal is to obtain the feature space depicted in Figure 4.12 which makes classification easier. The MRM is a computationally simple method. Using example values, the steps are described in Figure 4.15 and are outlined below:

1. LFP recordings from each monopolar channel are split into M training epochs with an equal number of PD and non-PD training examples. In Figure 4.15, $M = 128$ results in 64 epochs for both PD and non-PD.

2. Each epoch is divided into N-bands (features). For each feature, the sum of that feature for all training examples for both PD and non-PD cases are obtained. In Figure 4.15, $N = 5$ is selected with the power in each of the five bands for both PD and non-PD training examples shown.

3. For the summed features (calculated in step 2), the ratio of correspondingly indexed features for PD and non-PD are calculated. The smaller of the two is made the divisor. The division is indicated in red in Figure 4.15 with the obtained result shown in step 3 of Figure 4.15.

4. The ratios are arranged in descending order. This order shows the relative vari-

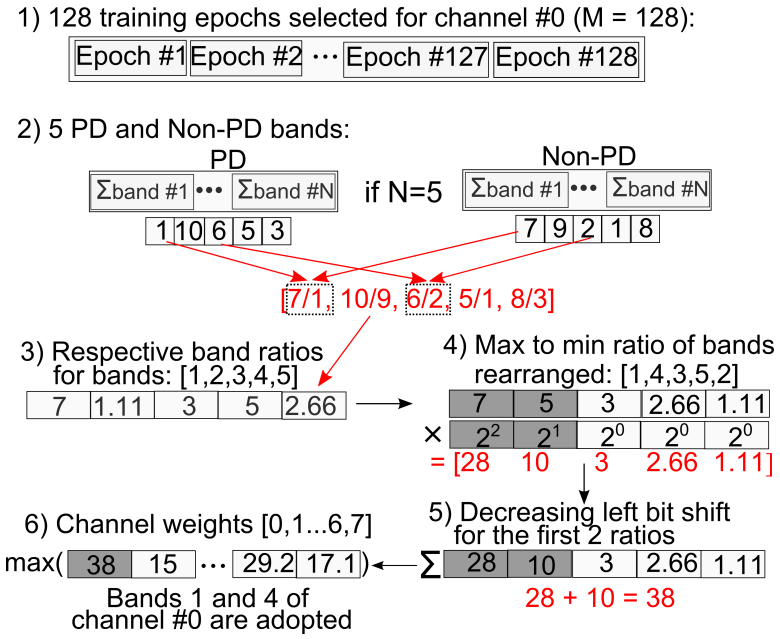


Figure 4.15: Detailing the MRM. This shows a situation where the channel and feature selection for MRM uses only the two prominent features in ranking the channels.

ation for each feature between PD and non-PD for the training examples – from the largest to the smallest.

5. The maximum ratio is multiplied by 4 and the second maximum ratio is multiplied by 2 and the rest are unchanged. If k-features are used for classification, the weighted sum of the first k-features gives the channel weight. In Figure 4.15, if weights based on two features are required, step 5 shows the channel weight in red which is 38. The channel weight gives a low computation approximation of the channel with the largest Euclidean norm between PD and non-PD clusters for the selected features. The approximate Euclidean norm rank is obtained using a modified version of [236].

6. Steps 1 to 5 are repeated for the rest of the channels. The channel with the largest weight is selected and recordings from that channel are used for classification until another training phase, after which the new highest ranked channel is adopted. Using the values in step 6 of Figure 4.15, the channel with the maximum weight is selected, which is channel #0 with a weight of 38.

4.4.3 Dynamic Pattern Classification

Dynamic classification uses a modified version of traditional pattern classifiers to accurately track the non-linearities in the extracted features. The classifiers must not be so simple that they are unable to distinguish between classes, yet not so complex as to over-fit the training data [85]. The best classifier will be selected based on a trade-off between computational complexity and performance. The dynamic pattern classifier steps through three orders of the traditional pattern classifiers to be evaluated. The first order classifier is used if it achieves a classification accuracy greater than 90% on

validation data, else a higher order classifier is invoked until the criteria is satisfied. If the criteria is not satisfied, the best performing classifier of the three is used. Below is a brief description of the evaluated pattern classifiers.

4.4.3.1 Dynamic k-Nearest Neighbour (k-NN)

k-NN uses a non-parametric method for classification. Amongst the various classes to distinguish between, it uses the predominant k closest samples in the feature space in classifying unlabeled points, where k is a natural number [10]. The function $f_{kNN}(\lambda)$ for k-NN is [10],

$$f_{kNN}(\lambda) = \sum_{n \epsilon N_k(\lambda)} \Upsilon_n \tag{4.27}$$

where λ is the test case, Υ_n are the labels for the closest training datasets, $N_k(\lambda)$ is the index of k-nearest neighbours of λ in the training set. Generally, k-NN does not require the normal iterative learning phase necessary in order to fit the training data to a classification model. The dynamic k-NN steps through 3-NN, 5-NN and 7-NN in that order. Odd nearest neighbours are employed for the dynamic classifier because there are two classes and simple majority voting is required.

4.4.3.2 Dynamic Logistic Regression (LR)

LR uses probability of class membership for predicting a test case. Considering a binary classification problem, with class membership $\Upsilon \epsilon \{0, 1\}$, 1 for PD cases and 0 for non-PD cases, $\lambda^{(i)}$ the extracted features for test case i, with a corresponding label $\Upsilon^{(i)}$. Then the logistic function $f_{LR}(\lambda)$ for classification in Equation 4.28 is [237],

$$f_{LR}(\lambda) = g(\theta^T \lambda) = \frac{1}{1 + e^{-\theta^T \lambda}} \tag{4.28}$$

which produces an $f_{LR}(\lambda)$ between 0 and 1, in order to predict the binary states. θ is a vector of threshold and weight parameters that is specific to a dataset. Using the logistic function $f_{LR}(\lambda)$, the two possible outcomes of PD detection can be represented as

$$\Upsilon = \begin{Bmatrix} 1 & if\ f_{LR}(\lambda) \geq 0.5 \\ 0 & if\ f_{LR}(\lambda) < 0.5 \end{Bmatrix} \tag{4.29}$$

where true PD detection is signified as 1, and false PD detection as 0. The polynomial order of the function $\theta^T \lambda$ determines the type of LR. For the dynamic LR, it steps through a linear function, third order and fifth order classifiers in that order. Functions having only odd degree polynomials are tested since the exponential parameter in the logistic function $f_{LR}(\lambda)$ in Equation 4.28, needs to be raised to a negative power.

4.4.3.3 Dynamic Support Vector Machine (SVM)

SVM uses the widest margin between differing states to discriminate. In Equation 4.30, the discriminating function $f_{SVM}(\lambda)$, used in classifying test cases is obtained using the training examples as [238],

$$f_{SVM}(\lambda) = \sum_i \Upsilon_i \rho_i K(\lambda_i, \lambda) + \omega \tag{4.30}$$

where λ_i are the support vectors and their labels Υ_i, λ is the test case, $K(\lambda_i, \lambda)$ is the kernel transformation, ρ_i is a weight vector and ω represents the classification threshold. To handle the complex nature of physiological signals, the decision function can be transformed for use with different kernels, notably the linear kernel, polynomial kernel and the radial basis function (RBF). This can be achieved by replacing the $K(\lambda_i, \lambda)$ in the kernel transformation, with the appropriate kernel function. For dynamic SVM classification, a linear kernel, quadratic kernel and RBF are used in that order.

4.5 Results

The results were tested in MATLAB on the described LFP test datasets.

4.5.1 Feature Space

A visualization of the feature space formed by the two features selected using MRM for dataset C is shown in Figure 4.16. The STFT shows a clearer separation (for PD, Trans and non-PD training examples) compared to the DWT, due to the use of a reduced frequency band by the STFT. A reduced band provides better confinement of the relevant frequencies and reduces sensitivity to outliers. The two methods are compared to identify the optimum performance in terms of detection accuracy and complexity. Generally, STFT is ideal for capturing sinusoidal features, and the DWT is ideal for detecting non-continuous frequencies. This was concluded in [233] using the Haar wavelet.

For toolbox based analysis of DWT (such as FieldTrip and EEGlab), Morlet wavelets are used. For toolbox based STFT, spectral smoothing is introduced using multi-

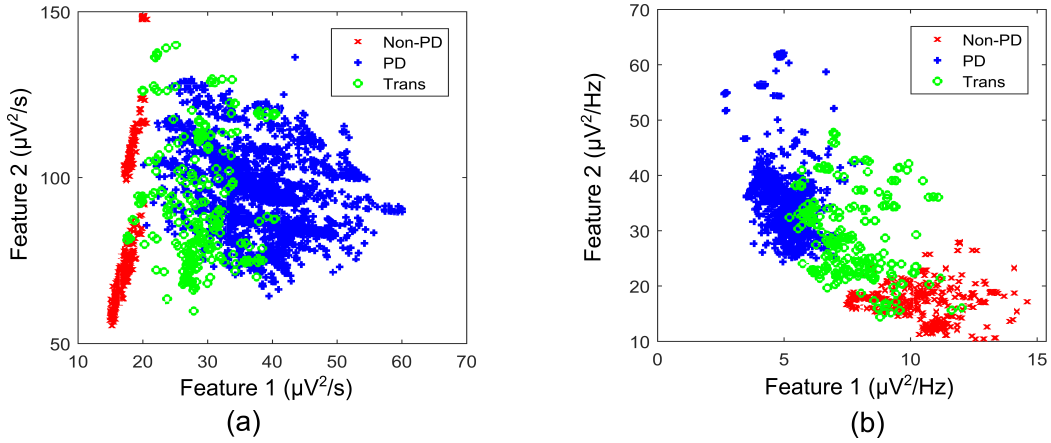


Figure 4.16: Feature space formed using MRM for the two prominent features of the selected channel of dataset C using: (a) DWT (32 - 49.5 Hz band power and 16 - 32 Hz band power) (b) STFT (21-26 Hz band power and 18- 23 Hz band power). LFP epochs at transition points (consisting of PD and non-PD of equal length) are labelled "Trans".

tapering. However, time-frequency analysis in this work was guided by considerations for hardware implementation, which are not fundamental for toolbox based analysis. For hardware implementation, the possible improvement in accuracy due to spectral smoothing, is not commensurate with the computational cost incurred. As demonstrated in Section 4.5.2 on accuracy, all the examined combinations achieve less than 10% classification error even without spectral smoothing. Both DWT (Haar wavelet) and STFT (without multi-tapering) have been shown as suitable in hardware-aware implementations for time-frequency analysis [211].

4.5.2 Accuracy and Complexity Measures

Various measures such as choice probabilities, classification error and F1-score were used in order to evaluate the detection methods. In each case the average result from 100 runs of Monte Carlo analysis was found to obtain the general trend.

4.5.2.1 Optimum Parameters

For the classification error averaged over all datasets, the goal was to obtain the minimum parameters (minimum number of features, minimum level of quantisation and minimum training examples) that resulted in 90% classification accuracy (10% classification error), to compute the complexity of each combination. Figure 4.17 presents the effect of varying the features used in classification. In Figure 4.17(a), the results for the combinations that use DWT for feature extraction are presented; the k-NN based algorithms (combinations) present the best performance, having a classification accuracy greater than 90% irrespective of the number of features used for classification. The second best performance was obtained by the SVM based classifiers, with the DWT-MRM-SVM having a superior performance compared to the DWT-PCA-SVM, even though both require at least two features to attain the 90% accuracy mark. The LR based algorithms behave in a similar way as the SVM based algorithms; however, they have a more gradual slope. In Figure 4.17(b), the combinations using STFT for feature extraction have nearly identical characteristics as those using DWT for feature extraction. That is, the k-NN based classifiers have the best performance, followed by the SVM based classifiers and then the LR-based classifiers. Like the DWT algorithms in Figure 4.17(a), the STFT algorithms in Figure 4.17(b) achieve an accuracy of 90% with two or less features.

With respect to the minimum training examples, Figure 4.18 presents the performance of the classifiers. For a classification accuracy of 90%, the SVM based algorithms require few training examples compared to the LR and k-NN based algorithms. In all three cases of Figure 4.18, algorithms using MRM for dimensionality reduction require many fewer training examples compared to their PCA counterparts. Another notable characteristic is that the combinations using STFT require fewer training examples compared to those using DWT. This may be attributable to the narrower frequency bands used in STFT, which makes it easier to discern patterns with few training examples – wider frequency bands like in DWT may be more susceptible to noise.

However, in Figure 4.19 which summarises the required quantization for the various

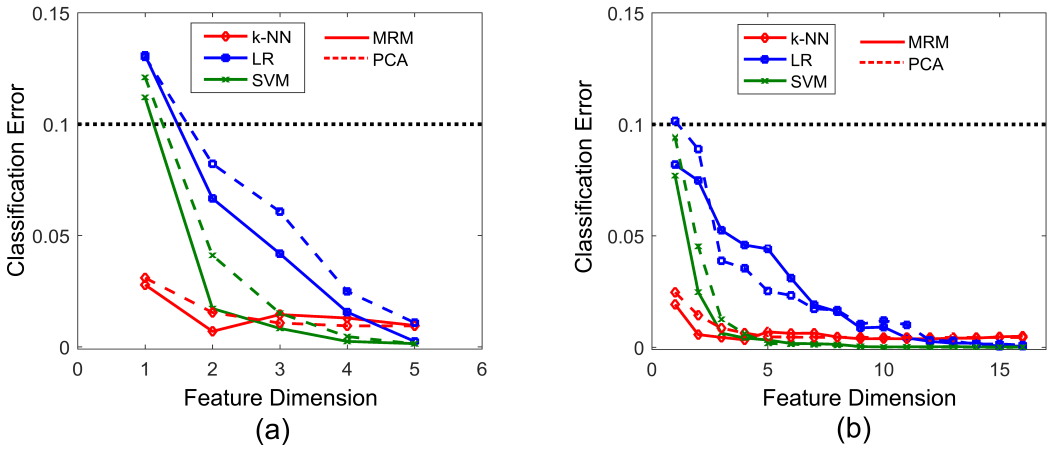


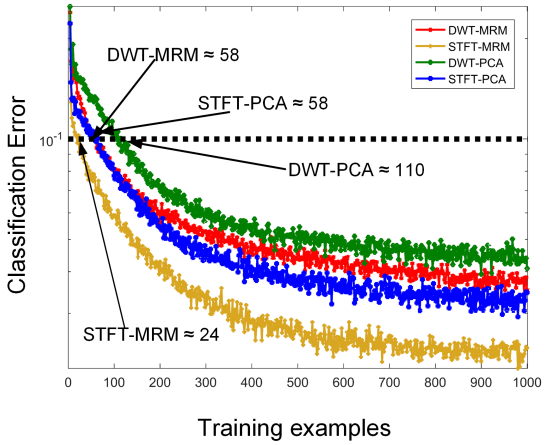
Figure 4.17: Effect of feature vector dimension on classification error using: (a) DWT (b) STFT. The plots with the dashed lines are those using PCA for DR, while those without dashes use MRM for DR. The goal is to determine the minimum number of features that achieve a classification accuracy of 90% (10% classification error).

algorithms The PCA requires fewer quantization bits than the MRM based algorithms. The PCA based algorithm requires fewer quantisation levels possibly because of the need for less detail in the number representation after PCA transformation. Training examples are less packed together after PCA transformation because PCA transforms features to a new feature space with a higher variance. The need for less detail after PCA transformation enables the use of fewer quantisation levels.

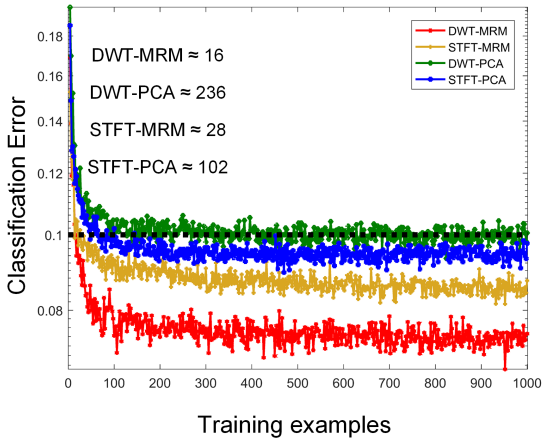
4.5.2.2 Compexity Measures

The minimum parameters needed to achieve a classification accuracy of 90% for the rest of the algorithms is summarised in Table 4.5. It can be seen that the k-NN and LR based algorithms have a higher median number of training examples, compared to the SVM. This is because the k-NN and LR are population dependent algorithms that extrapolate properties more accurately when larger training sets are used. In Table 4.5, it can be seen that 7 combinations require only one feature to achieve 90% classification accuracy. Regarding the quantisation, the median is 6 bits resolution. Table 4.5 also summarises the parameters that achieve 90% classification accuracy if all features are used for dynamic detection. From Table 4.5, when the maximum number of features are used, the required training examples and quantisation needed to achieve 90% classification accuracy are reduced. However, for the STFT-MRM combinations (STFT-MRM-KNN, STFT-MRM-KNN and STFT-MRM-SVM) using all features presents a lower NOP for 90% classification accuracy compared to its low feature alternative. This may be because computing STFT and extracting power from a single band (single feature) requires slightly less computation than extracting power from all the allocated bands (16 in this case). However, the extra computation incurred is offset by the fewer training examples and quantisation levels required when all the bands (features) are used. Mathematically the NOP calculation is given by,

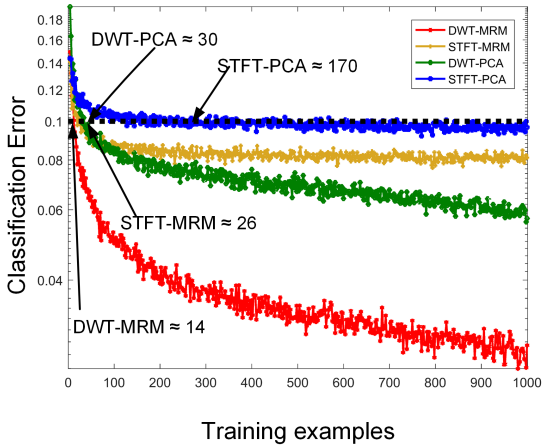
$$NOP = \frac{Training\ NOP}{86400} + Operating\ NOP \tag{4.31}$$



(a)

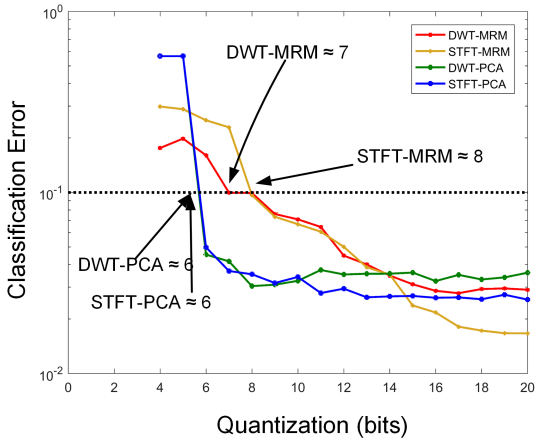


(b)

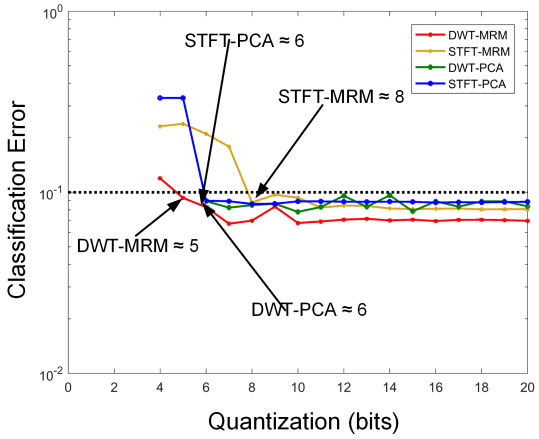


(c)

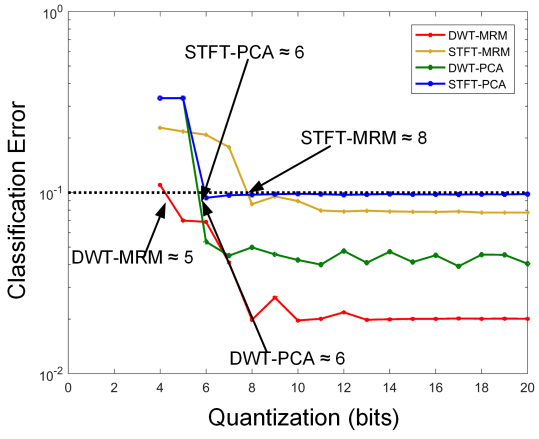
Figure 4.18: Optimum values of training examples to achieve 90% classification accuracy for classification using: (a) k-NN (b) LR (c) SVM.



(a)



(b)



(c)

Figure 4.19: Optimum quantisation (bits) to achieve 90% classification accuracy for classification using: (a) k-NN (b) LR (c) SVM..

The training NOP is divided by 86400 because it is assumed that training occurs once in a day (86,400 seconds), and the patient state is updated every second during normal operation.

Table 4.5: Combinations of algorithms and the level of complexity needed to achieve 90% classification accuracy.

Combination	Number of features	Quantization (bits)	Number of training examples	Complexity	
				NOP ($\times 10^5$)	Area (mm^2)
DWT-MRM-KNN	1	7	58	0.26	0.84
	5	5	14	0.61	1.48
DWT-PCA-KNN	1	6	110	0.24	0.81
STFT-MRM-KNN	1	8	24	0.90	2.12
	16	4	22	0.36	1.15
STFT-PCA-KNN	1	6	58	0.56	1.40
DWT-MRM-LR	2	5	16	0.24	1.05
	5	8	8	1.51	3.98
DWT-PCA-LR	2	6	236	0.39	1.84
STFT-MRM-LR	1	8	28	0.89	2.70
	16	4	20	0.31	1.26
STFT-PCA-LR	2	6	102	0.56	1.96
DWT-MRM-SVM	2	5	14	0.24	1.05
	5	4	8	0.39	1.25
DWT-PCA-SVM	2	6	30	0.37	1.45
STFT-MRM-SVM	1	8	26	0.89	2.70
	16	4	8	0.29	1.11
STFT-PCA-SVM	1	6	170	0.57	1.94

4.5.2.3 Choice Probability and F1-score

Using the optimum combination for quantization, features and training examples that achieved 90% classification accuracy (as summarised in Table 4.5), Figure 4.20 presents the performance of the algorithms in terms of their F1-score and choice probability. For the classifiers, it is clear that the k-NN based algorithms present the best performance (having the least error cost), followed by the SVM-based and then the LR-based. The STFT-MRM-KNN has the least error cost of 0.0075 [Figure 4.20 (b)], while the STFT-PCA-SVM has the maximum error cost with 0.0778 [Figure 4.20 (d)].

4.5.3 Combined Cost

The combined cost represents the cost incurred by a given combination. This is summarised in Table 4.6. Figure 4.21 shows a plot of the normalised error cost versus the normalised complexity cost. An ideal detection algorithm is required to have a

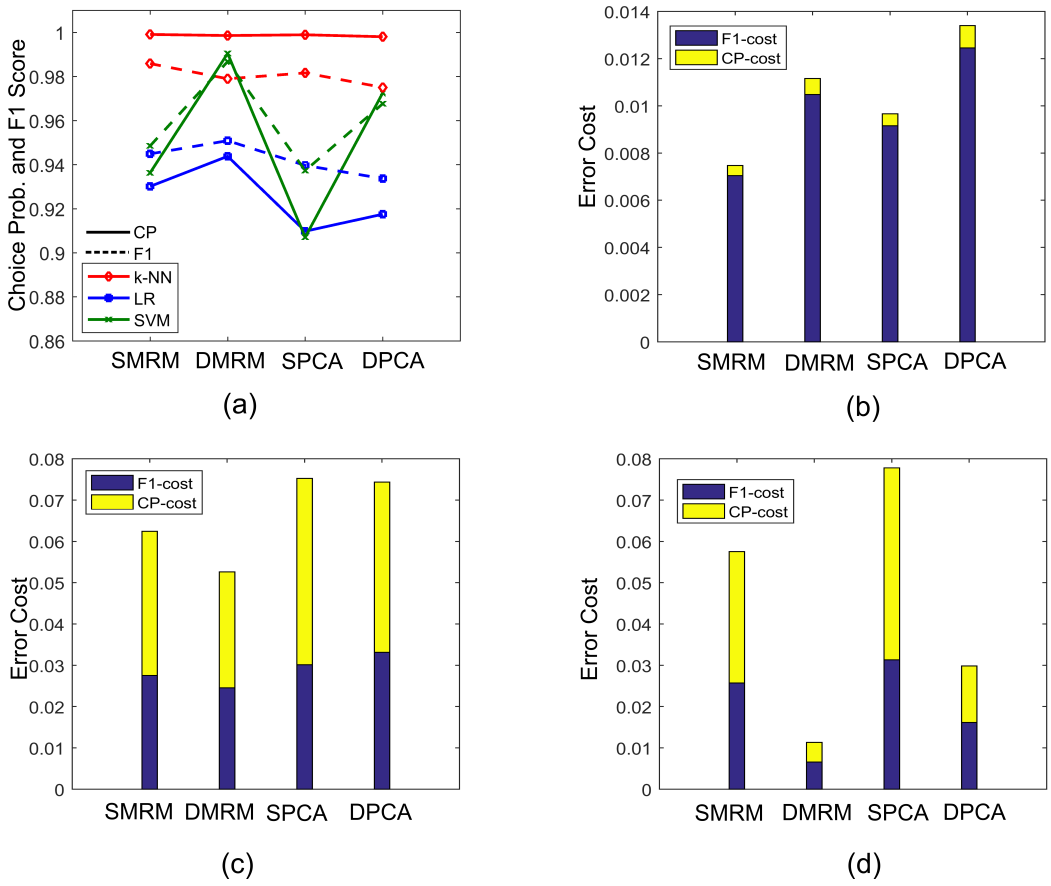


Figure 4.20: Plot of choice probability and F1-score (SMRM is STFT-MRM, DMRM is DWT-MRM, SPCA is STFT-PCA and DPCA is DWT-PCA). (a) The plot of choice probability and F1-score for different combination of algorithms; (b) Plot of error cost for algorithms using k-NN for classification; (c) Plot of error cost for algorithms using LR for classification; (d) Plot of error cost for algorithms using SVM classification..

combined cost at the origin of Figure 4.21. The error cost is obtained at 10% classification error and consists of costs resulting from a low average choice probability, and a low average F1-score. The choice probability is included because it shows how well a detector ranks PD cases compared to non-PD cases. The F1-score is included because it shows how good the precision and sensitivity of the detector is. These two factors are not apparent when only classification error is used in assessing the error cost of a detector. They are added to ensure that the error cost is robust enough to cover all accuracy measures. For the complexity cost, the NOP and estimated area are used as the measures. In Figure 4.21, the detector with the low costs are indicated by blue markers, and are the DWT-MRM-KNN, DWT-MRM-SVM and DWT-PCA-KNN. There are six medium cost detectors indicated by green markers. The high cost detectors are indicated by red markers and are the STFT-PCA-SVM, DWT-PCA-LR and STFT-PCA-LR. The algorithms using DWT for feature extraction, are mainly closer to the origin. Hence the DWT based algorithms have the optimal trade-off between complexity and accuracy. For dimensionality reduction, the MRM based algorithms have the lowest cost while for classification the k-NN based algorithms have the lowest cost.

Table 4.6: Summarising the error and complexity costs of different combination of algorithms.

Combination	Error Cost	Complexity Cost
DWT-MRM-KNN	0.0112	0.4424
DWT-PCA-KNN	0.0134	0.4172
STFT-MRM-KNN	0.0075	0.6092
STFT-PCA-KNN	0.0097	0.8484
DWT-MRM-LR	0.0526	0.4784
DWT-PCA-LR	0.0743	0.8115
STFT-MRM-LR	0.0624	0.5934
STFT-PCA-LR	0.0752	0.9912
DWT-MRM-SVM	0.0113	0.4784
DWT-PCA-SVM	0.0299	0.6945
STFT-MRM-SVM	0.0575	0.5375
STFT-PCA-SVM	0.0778	0.9949

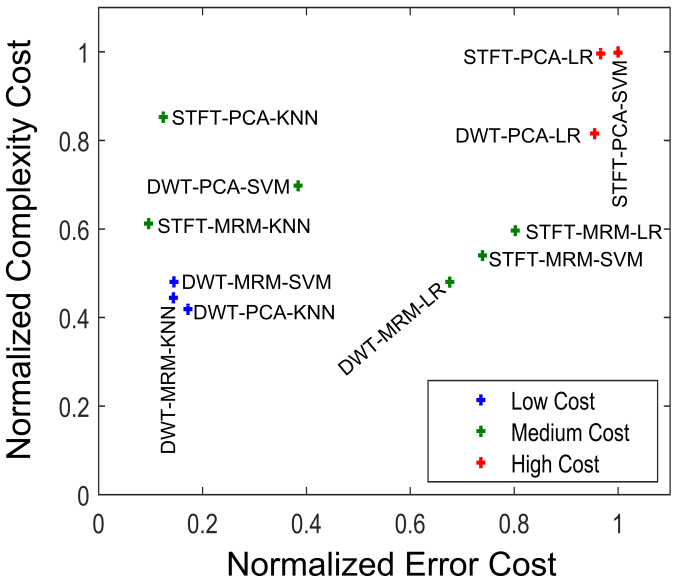


Figure 4.21: The combined cost for different combinations of algorithms. Both axes are normalized such that the maximum cost in each case is equal to 1.

Thus, the combination closest to the origin is the DWT-MRM-KNN, which represents the optimal trade-off between accuracy and complexity; while the worst performing algorithm is the STFT-PCA- SVM.

4.6 Discussion

4.6.1 Semi-Synthetic Datasets

Semi-synthetic data was generated by taking advantage of statistical properties in the data to fit ARMA models. Moreover, for every single epoch, a range of Monte Carlo forecasts were simulated based on the underlying variations. To the best of our knowledge, there is no literature on semi-synthetic LFP data generation using real-life recordings. However, for other signals such as spikes and ECG, there is extensive literature.

Using statistical parameters like the ones used in defining spike [239] or ECG [240] activity could be misleading; because they have a unique morphology which can be varied using simple measures, such as the amplitude or shape. On the other hand, LFPs result from the activity of a localised population of sources; hence using statistics that are not population-based corrupts and destroys the signal fidelity. In this chapter some properties were exploited, among which are:

1. *Statistical Similarity*: To measure the sensitivity of the algorithms to all forms of PD and non-PD variations, it was ensured that statistically dissimilar epochs (measured using the normalised cross-correlation) were attached together during periods of long PD or long non-PD synthesis. This introduced some randomness by attaching epochs with varying similarity, which enabled the assessment of the algorithms' robustness to instantaneous changes within the same patient state.

2. *Dispersion and Random Permutation Sampling*: Dispersion was introduced using ARMA models to forecast a range of Monte Carlo variants for each epoch. This was to create a large diversity pool that overcomes sampling bias (since sampling bias can lead to poor generalizability). Epoch selection using random permutation sampling was applied to avoid data leakage which can cause over generalizability.

3. *Poisson Distribution Defined PD and non-PD Duration*: Signal length was defined using a Poisson distribution to make PD and non-PD episodes pseudo-random. This ensured that a "randomly guessing" algorithm that changes state prediction based on a predefined pattern is flagged because of the pseudo-random PD and non-PD periods.

4.6.2 Spectral Bands and Maximum Ratio Method

In the normalized autospectra shown in Figure 4.22 the largest variation between the PD and non-PD autospectra is between $10-25$ Hz, which mostly lies in the beta band; $13-30$ Hz. The gamma band (>30 Hz) shows little relative activity. However, in Table 4.7 where the two frequency bands with the most activity for each of the nine datasets are summarised, it can be seen that the gamma band is not strictly without activity. Datasets A, F and G have their most pronounced variation in the gamma bands (level 1 detail coefficients) when DWT is used for feature extraction. While using STFT, it was only visible for dataset G. The little activity in gamma bands is corroborated by [122], in which there are occasions where there is little activity, while at other times they have the most pronounced activity.
The frequency bands for the STFT and DWT in Table 4.7 appear to be unexpectedly different. This is due to different sized frequency bands. For instance, in the dataset for patient D, the $0.5-5$ Hz band presents the most pronounced variation for STFT, while the $4-8$ Hz band provides the most variation for DWT. This shows that most of the power is within the $4-5$ Hz band, giving rise to the dominance of the bands containing these frequencies. The difference between the STFT and DWT in other cases

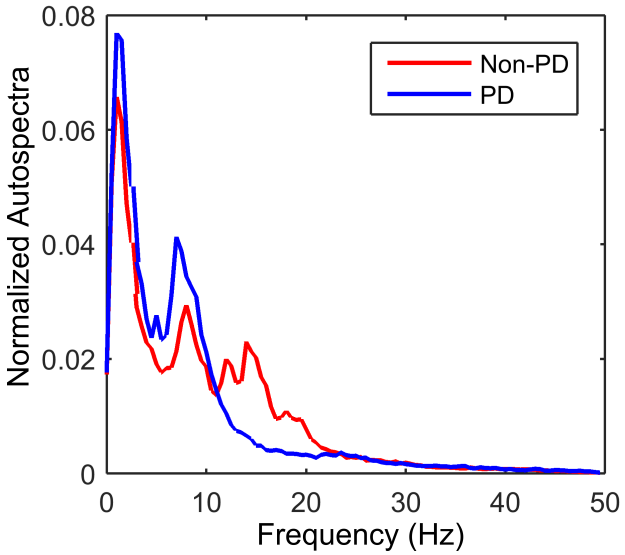


Figure 4.22: Average normalized autospectra for all the datasets. The PD and non-PD plots are normalized such that the total autospectra sum in each case is equal to 1.

Table 4.7: Summarising the two frequency bands with the most pronounced variation.

Dataset	STFT		DWT	
	Maximum variation band (Hz)	2^{nd} maximum variation band (Hz)	Maximum variation band (Hz)	2^{nd} maximum variation band (Hz)
A	6-11	9-14	32-49.5	4-8
B	12-17	15-20	16-32	8-16
C	21-26	18-23	32-49.5	16-32
D	0.5-5	6-11	4-8	8-16
E	12-17	15-20	8-16	16-32
F	18-23	15-20	32-49.5	16-32
G	45-49.5	42-47	32-49.5	16-32
H	36-41	33-38	0.5-4	4-8
I	18-23	21-26	16-32	32-49.5

can be explained because the most pronounced variation is dependent on the relative power between PD and non-PD. That is, there are cases in which a PD band can have the greatest absolute activity compared to other PD bands. But when compared to its corresponding non-PD band, it may not have the most pronounced variation. This was the case for datasets A and H. As a result of the varying frequency spread present, particularly in the DWT, only the activity of the much lower bands of the DWT should be expected to show some semblance in behaviour to the STFT. These are the cases where the frequency spread is quite close and at the higher frequencies, they cannot be compared because a single DWT band covers more than five STFT bands. In Figure 4.23 , it is shown that the MRM algorithm can detect the bands with the most pronounced variation even though the training examples are a small fraction of the entire population. In Figure 4.23 (a), it accurately detects the band with the most pronounced variation for all 100 runs while in Figure 4.23 (b) it accurately detects

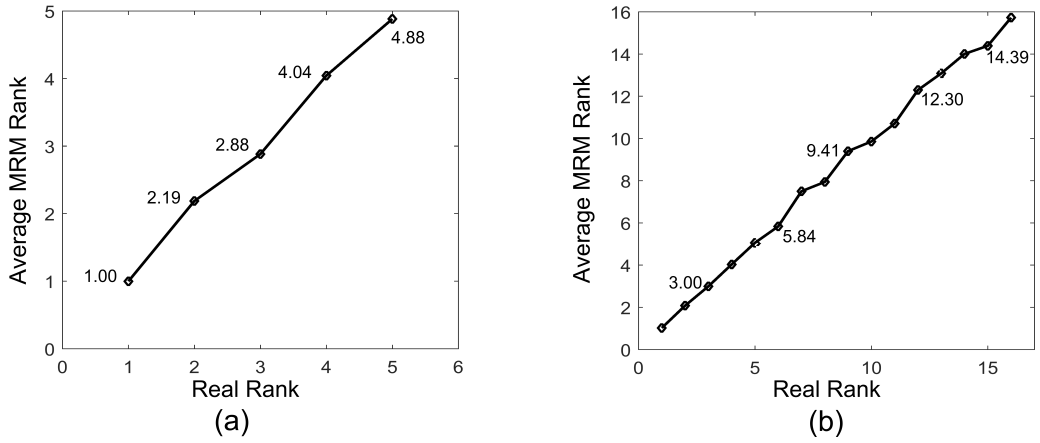


Figure 4.23: Shows the mean plot of how the MRM algorithm ranks the frequency bands using: (a) DWT for feature extraction (b) STFT for feature extraction. The rank is from the band with the most pronounced variation (ranked 1) to the least pronounced (ranked 5 for DWT and 16 for STFT).

the three prominent bands for all runs. Most of the average ranks approximate to the real rank showing that most of the time, it ranks the bands in their right order. This demonstrates that the training examples used have enough diversity present in the entire signal. The MRM algorithm also assists in channel selection by approximately determining the channel with the largest Euclidean norm between PD and non-PD clusters. Hence, instead of using all eight monopolar channels for detection, only the channel with the most pronounced variation is selected. This makes classification less computationally intensive.

4.6.3 Dynamic Detection

The dynamic schemes are used to obtain a classifier that is compatible with the dynamic FE and at the same time achieves the right balance between complexity and classification accuracy. The dynamic classifier operates in two phases – concurrent detection and training, and detection only. Compatibility between dynamic FE and dynamic pattern classification is determined periodically during the concurrent detection and training phase. The dynamic detection operates in real-time since both phases involve detection. Determining compatibility is necessary because high order classifiers may sometimes provide inferior performance compared to low order classifiers, making the use of high order classifiers redundant. A classic example of this is shown in Figure 4.24(a), which uses the STFT-MRM-KNN algorithm. Apart from the region where few features (one to four features) are used in which there is a struggle for dominance, the 3-NN completely dominates when five or more features are used. Thus, making the higher order classifiers (5-NN and 7-NN) redundant. Nevertheless, the benefits of dynamic detection are obvious since a mean reduction of 0.019% in classification error is obtained between the dynamic classifier and the 3-NN. This is because different kernels are selected for different datasets (A–I). In the long term, these slight differences of 0.019% between the dynamic classifier and the best performing static classifier can cumulatively lead to more beneficial effects in therapy since there is a higher likelihood

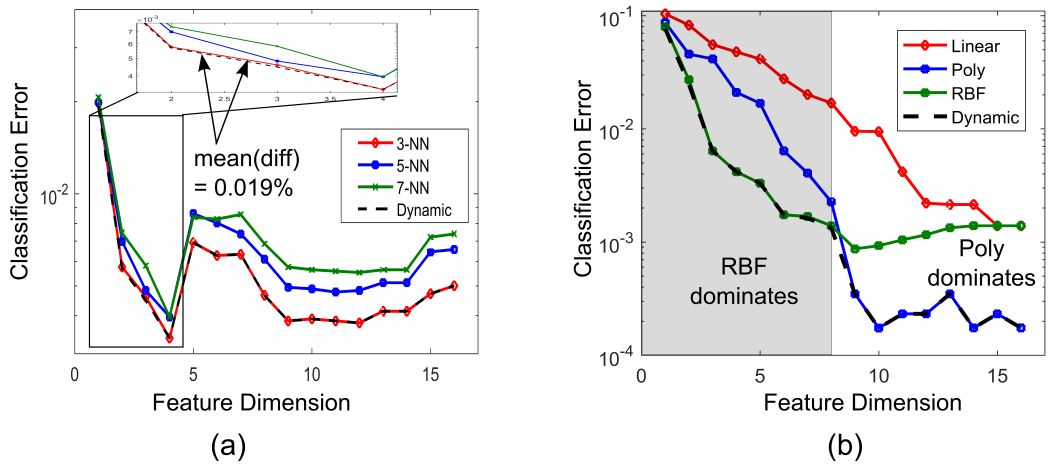


Figure 4.24: Comparison of the classification error for static and dynamic detection schemes: (a) using the STFT-MRM-KNN detector; (b) using the STFT-MRM-SVM detector.

of detection. Conversely, in Figure 4.24(b), which depicts the result for the STFT-MRM-SVM algorithm, there are two regimes of operation: when eight features or less are used for classification, the RBF, which is the highest order kernel dominates. On the other hand, when more than eight features are used the lower order poly kernel dominates, similar to Figure 4.24(a). This shows that dynamic detection takes advantage of the best operating regimes of various kernels to produce performance superior to those of static classifiers as is shown in Figure 4.24. As a result, the three different classifiers in the dynamic scheme are complementary.

4.7 Chapter Summary

The major contributions presented in this chapter are:

a) We presented dynamic detection as a way to overcome the unpredictable nature of PD pathophysiology.

b) It was demonstrated that the change in power characteristics of LFP signals, can be tracked using a combination of algorithms consisting of FE, DR and dynamic pattern classification.

c) For hardware efficient implementation of dynamic PD detection, a subset of algorithms consisting of the aforementioned stages were evaluated based on customised performance metrics. The metrics were chosen to reflect detection accuracy and computational complexity trade-offs.

d) The combination with the the best trade-off between computational complexity and detection accuracy, consisted of DWT for feature extraction, MRM for DR and dynamic k-NN for classification. The results presented show that on-chip (online) PD detection is possible.

e) To enable a fair and balanced evaluation of the algorithms, semi-synthetic LFP

datasets were generated from real LFP recordings. Also, the paucity of validation datasets also motivated the semi-synthetic data generation. Semi-synthetic LFP generation was achieved using statistical properties of the original LFP signals. For example, time series forecasting using ARMA models were used to create semi-synthetic LFP signals. To ensure that the synthesised LFP signals mimicked the non-linear progression in PD, PD events were defined in a pseudo-random manner using Poisson distributions. And, to avoid training data leakage into test data, random permutation sampling was used for LFP epoch selection and concatenation.

f) The MRM, inspired by Fisher's separability criterion, was presented as a computationally efficient implementation of DR. To facilitate fully implantable aDBS, computationally efficient implementations of each of the stages in the processing chain are required.

g) In addition to PD detection, the algorithm could be extended to other applications that require on-demand DBS for efficient modulation of therapy.

Chapter 5

Complementary Detection for Hardware-Efficient Monitoring of Disease Progress in Parkinsonian Patients

5.1 Introduction

PD has complex mechanisms [241], and to optimize therapy, a better understanding of its dynamics is required. Currently, the standard for diagnosing and monitoring parkinsonian progression in patients is through visual feedbacks from patients [22]. These may be insufficient since it is only monitored during follow-up visits. Research in disease monitoring has ranged from using mobile devices that have short message service (SMS), web-based applications and Bluetooth capability to measure the frequency of symptom onset so that medical interventions could be delivered or better diagnosis could be made [188]. These systems could be implemented on software application running on the patient's commercial smartphone and connected to the clinician's information systems [242]. Rossi et al. [243] proposed and developed a webbased system for collecting clinical and neurophysiological data, the WebBioBank, specifically created for DBS management, that also can be connected to the patient's mobile apps and can safely be used for web-based tele-monitoring and caregiver support [188]. Also, disease monitoring can be used to provide a more refined therapy and for biomarker selection based on patient data collected. The power required to transmit data in neural signal processing systems is shown to dominate that for recording and data conversion [244]. As such offline processing based on transmitting raw time series data as suggested in [188, 243], is an inefficient approach. Hence, based on the power and bandwidth constraints involved in continuously sending neural signals, it will be more resource efficient to periodically send patient progress as state estimates after on-site and online analysis. Such an integrated platform for on-site and online analysis and monitoring of PD signals is still unavailable.

For on-site and online analysis, there is a need to develop miniaturised real-time platforms that could monitor disease progress. This specialised hardware platforms will

facilitate mobile diagnostics for better disease management. Ambulatory platforms for PD diagnostics could provide more in-depth information and reduce the number of face-to-face visits required to optimize therapy. In PD monitoring, the aim is to provide long-term monitoring of the patient's condition for clinicians to better understanding the symptoms such that therapy could be tailored to patient needs. There are two major components to personalised health monitoring, namely; disease diagnosis and disease management. The focus here will be on disease diagnosis. This work presents an interface that can process LFP signals such that Parkinsonian states are communicated and logged onto an external platform. The processing chain is shown in Figure 5.1. The PD detection processor detects PD events and sends these events over a communication interface (instead of sending raw neural data) to a PD event log. This reduces the amount of data sent over the communication link, which in turn reduces communication energy as well as bandwidth requirements of the entire system.

The platform was implemented on FPGA. The benefits of FPGA are that they lie midway between micro-controllers (high flexibility and low speed) and ASIC (low flexibility and high speed) on the flexibility and speed spectrum. Also, the relative ease with which they can be translated into ASICs as well as programming languages for micro-controllers is an added advantage. The long-term plan is to implement fully implantable systems that adapt DBS therapy. The first step towards this is the implementation of hardware platforms for monitoring PD progress.

The chapter is organized as follows. Section 5.2 details the major functional units of the PD detection processor. Section 5.3 describes the FPGA prototyping platform for the processor. The obtained results are described in Section 5.4. Section 5.5 discusses the prospects and challenges of personalised monitoring. The major points addressed in the chapter are summarised in Section 5.6.

5.2 System Overview

For detection algorithms, the more complex algorithms mostly provide the best performance. As such, the choices in choosing algorithms along the processing chain is guided by efficacy and complexity measures. Fully implantable hardware is expected to operate under high-reliability requirements and strict power-density regimes. For us to strike the optimal balance between efficacy and complexity, algorithmic and hardware optimisations were adopted for each stage on the processing chain. The PD detection algorithm performs feature extraction, feature selection and classification. Hardware efficient and implementable realisations of each of the listed stages are adopted for the

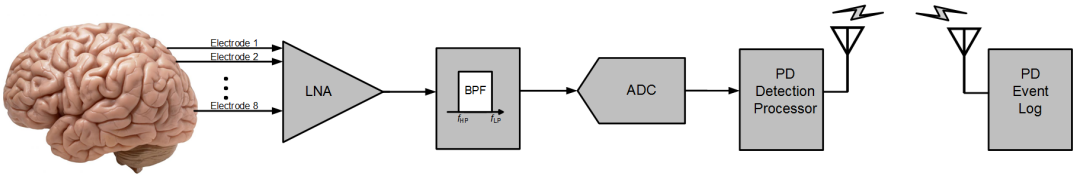


Figure 5.1: The functional block diagram of the PD monitoring platform.

full system. Figure 5.2 shows the top-level block diagram of the PD detection processor. The PD detection processor consists of five major functional units: feature extraction for training, feature extraction for detection, processor finite state machine (FSM), memory banks and k-NN unit. The PD detection processor has two operating regimes: concurrent training and detection, and detection only. During concurrent detection and training, time multiplexing is used to acquire training data from each of the recording channels. This is done through the feature extraction for training unit, which sends the extracted features to the memory banks for storing training data. The training examples stored are used to train the MRM (for feature and channel selection) as well as the k-NN classifier. In order to avoid interrupting PD detection, two feature extraction unit were used: the first serves training purposes and the second for detection. All these units are coordinated by the 'PD detection processor FSM'. The PD detection processor has four input clocks: *clk1* operates at 128 Hz and controls real-time acquisition of LFP data; the second is *clk2* which operates at 1 Hz and controls feature extraction. Features are computed from 256 samples/channel, with 50% overlap in samples. Five features are computed from the buffered 256 samples. Feature extraction is necessary because directly using time-series data as input to classifiers for detection may be impractical for real-time detection as these may increase the dynamic power of the classifier. This necessitated the need for two preliminary processing stages before classification, which are: feature extraction and feature selection. The MRM is used for feature and channel selection and it has two operating clocks: *clk1* and *clk3*. MRM training is conducted sequentially per channel and is controlled by *clk3*, whose frequency varies for various implementations of the PD detection processor. It varies between 1/81 Hz to 1/756 Hz depending on the implementation (the various implementations are discussed in later sections). A major consideration that drove the exploration of various architectures of the PD detection processor was to obtain a balance between performance and hardware resources utilisation. The final clock is *clk4*, which is only used by the 'PD detection processor FSM' to trigger a new training cycle; *clk4* is made 200 times slower than *clk3*. The following section describes the functional units in more detail.

5.2.1 Feature Extraction

Feature extraction is the first stage of neural data manipulation. During feature extraction, neural signals are transformed into features that separate non-PD from PD events. Another goal of FE is to extract as few useful information as possible to minimise the computational complexity of the detection process. A common algorithm that is widely used for time-frequency analysis of neural signals is the DWT. To extract useful information, DWT decomposes a signal into different levels based on frequency content. DWT is suitable for FE because the decomposition into different levels enhances the SNR of the neural signals; which facilitates the identification of PD and non-PD events. Examples of DWT algorithms are Morlet wavelet, Haar wavelet, and other Daubechies wavelets. The Haar wavelet has been commonly used due to its favourable balance between complexity of hardware implementation and detection performance.

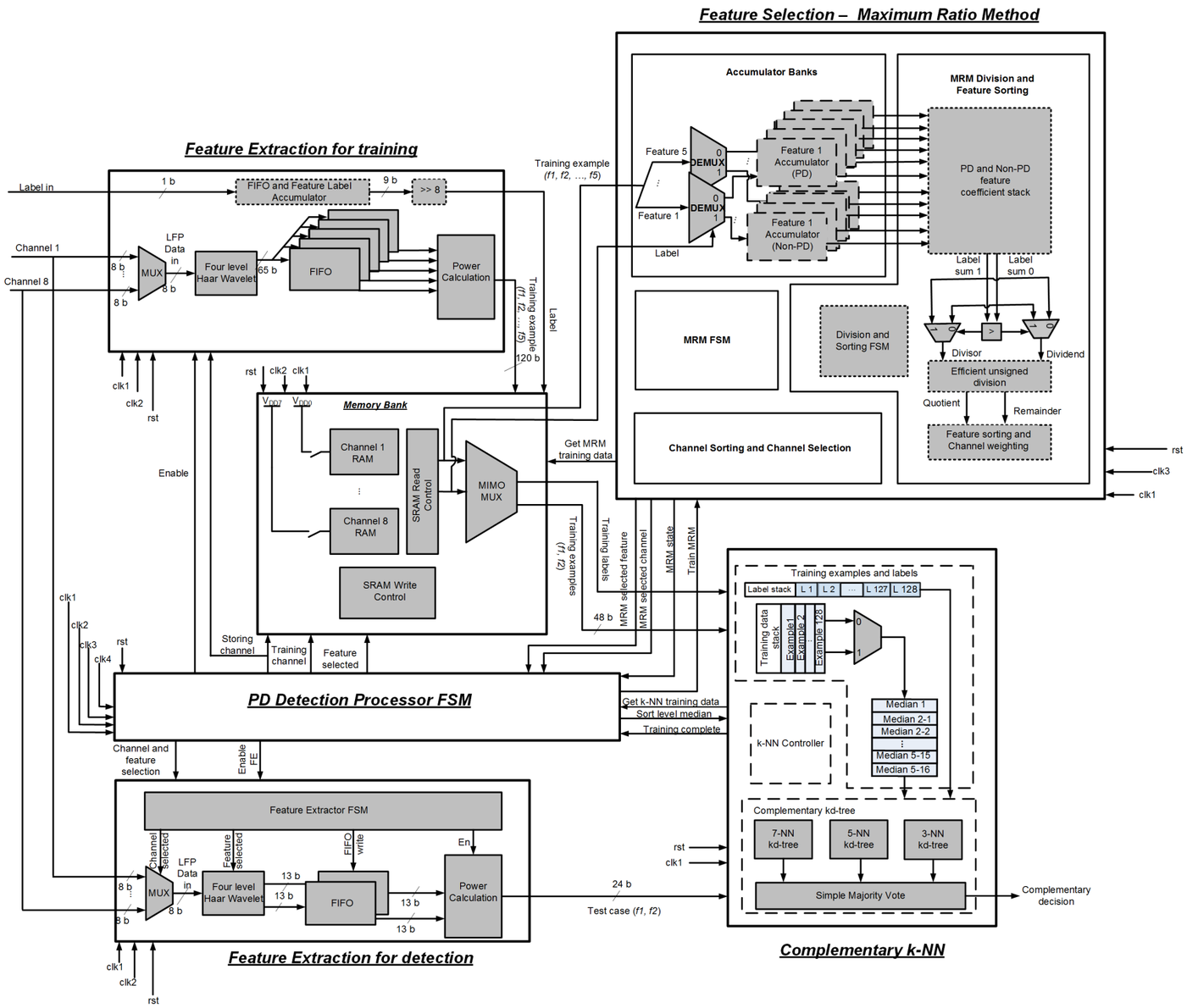


Figure 5.2: The functional units of the PD detection processor.

The Haar wavelet is ideal for capturing non-continuous frequencies [233]. In biomedical applications, the Morlet wavelet transform is widely used to discriminate abnormal behaviour due to its sensitivity to discontinuities like in Haar wavelet and its ability to offer multiple dimensions to time related changes. This advantage is leveraged upon in toolbox based analysis of DWT and it is used in applications like the detection of abnormal heartbeat behaviour. Since, time-frequency analysis in this work was guided by considerations for hardware implementation, which are not fundamental for toolbox based analysis. For the hardware implementation, the possible improvement in accuracy due to Morlet wavelet, may not be commensurate with the computational cost incurred. As demonstrated in Chapter 4, the Haar wavelet implementation presents high accuracy measures. Also, Haar-wavelets have been shown as suitable in hardware-aware implementations for time-frequency analysis [211]. An approach using four-level decomposition was adopted because it separated the features into desired brain wave bands as follows:

- Gamma band is greater than 30 Hz and the level 1 detail coefficients produce LFP characteristics between $32 - 49.5$ Hz. The input LFP signal is band-pass filtered between $0.5 - 49.5$ Hz, and then downsampled to fs $= 128$ Hz such that the maximum frequency is 64 Hz. A frequency domain demonstration of this is shown in Figure 5.3.
- Beta band activity is between $13-30$ Hz and Level 2 detail coefficients characterise LFP activity into frequencies between $16 - 32$ Hz as can be seen in Figure 5.3.
- Alpha band activity is $8 - 12$ Hz and level 3 detail coefficients are representative of LFP activity of frequencies between $8 - 16$ Hz.
- Theta band activity is between $3 - 8$ Hz and level 4 detail coefficients of the LFP activity are between $4 - 8$ Hz.
- Delta band is between $0.5 - 3$ Hz and the level 4 approximation coefficients obtained represented LFP activity between $0.5 - 4$ Hz.

Each level of decomposition results in down-sampling by two at each successive level, after four-level decomposition, further decomposition levels may not be desired because it will result in reduced frequency bands which may contain little or no relevant information.

5.2.1.1 Hardware Implementation

The two operating phases of the processor FSM are detection, and concurrent detection and training. During concurrent detection and training, a separate feature extraction

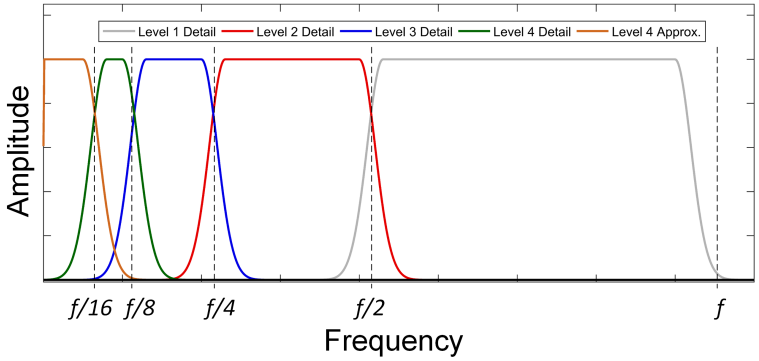


Figure 5.3: A frequency domain illustration of four-level wavelet decomposition.

stage is used to store training data from each of the channels while the other is used for detection. This is why there are two feature extraction units. Two feature extraction units are required so that real-time detection is not interrupted during training. One of the feature extraction units is always active and is used for detection, while the other is only active during training. To extract the five features in hardware, three major blocks are required:

1. A four-level Haar DWT block that computes the wavelet coefficients at each decomposition levels.

2. A power calculation block that computes the features at each level based on their corresponding coefficients.

3. And a first-in, first-out (FIFO) memory block to synchronise the four-level Haar-wavelet block with power calculation block.

Figure 5.4 (a) and (b) shows the structure of the two implementations of the feature extraction units. The four-level Haar wavelet block is synchronised to operate at the same sampling rate as the input data ($clk1 = 128\ Hz$) because the system is designed to adopt real-time PD detection. However, features are updated every one second. Thus, the power calculation block is synchronised to operate at $clk2 = 1\ Hz$. The power features are obtained using 2-second window (256 samples) of LFP signals with 50% overlap between windows. This means features are updated every second. The average power of the coefficients in each level can be obtained as follows,

$$P_i = \frac{1}{N_i} \sum_{n=0}^{N_i-1} \parallel x_i[n] \parallel^2 \quad . \tag{5.1}$$

where N_i is the number of DWT coefficients $x_i[n]$ at level i, with n representing the index for each sample. N_i is 128 for level-1 detail coefficients, 64 for level-2 detail coefficients, 32 for level-3 detail coefficients and 16 for level four coefficients. The four-level Haar wavelet block is synchronised to the power calculation block using a FIFO memory. The FIFO block is a dual-port RAM consisting of a memory and controller block. It also has separate read and write pointers that are used for controlling reading

and writing operations in the feature extraction unit. Below is a brief description of the two implementations of the feature extraction units.

Feature Extraction for Training

The feature extraction unit in Figure 5.4 (a) interfaces with the memory bank. During the processor FSM training phase, the feature extraction for training collects LFP training data from the multiple input channels through time-multiplexing. Data from all the channels use a common data path through the multiplexer. The active channel is controlled by input signal from the processor FSM. All coefficients from the four-level Haar wavelet are transferred to the power calculation block through a five-level FIFO corresponding to coefficients from each level. The FIFO for level 1 detail coefficients is made up of 128 memory locations, level-2 detail uses 64 memory locations and so on; with level 4 approximation coefficients consisting of 16 memory locations. Besides, the

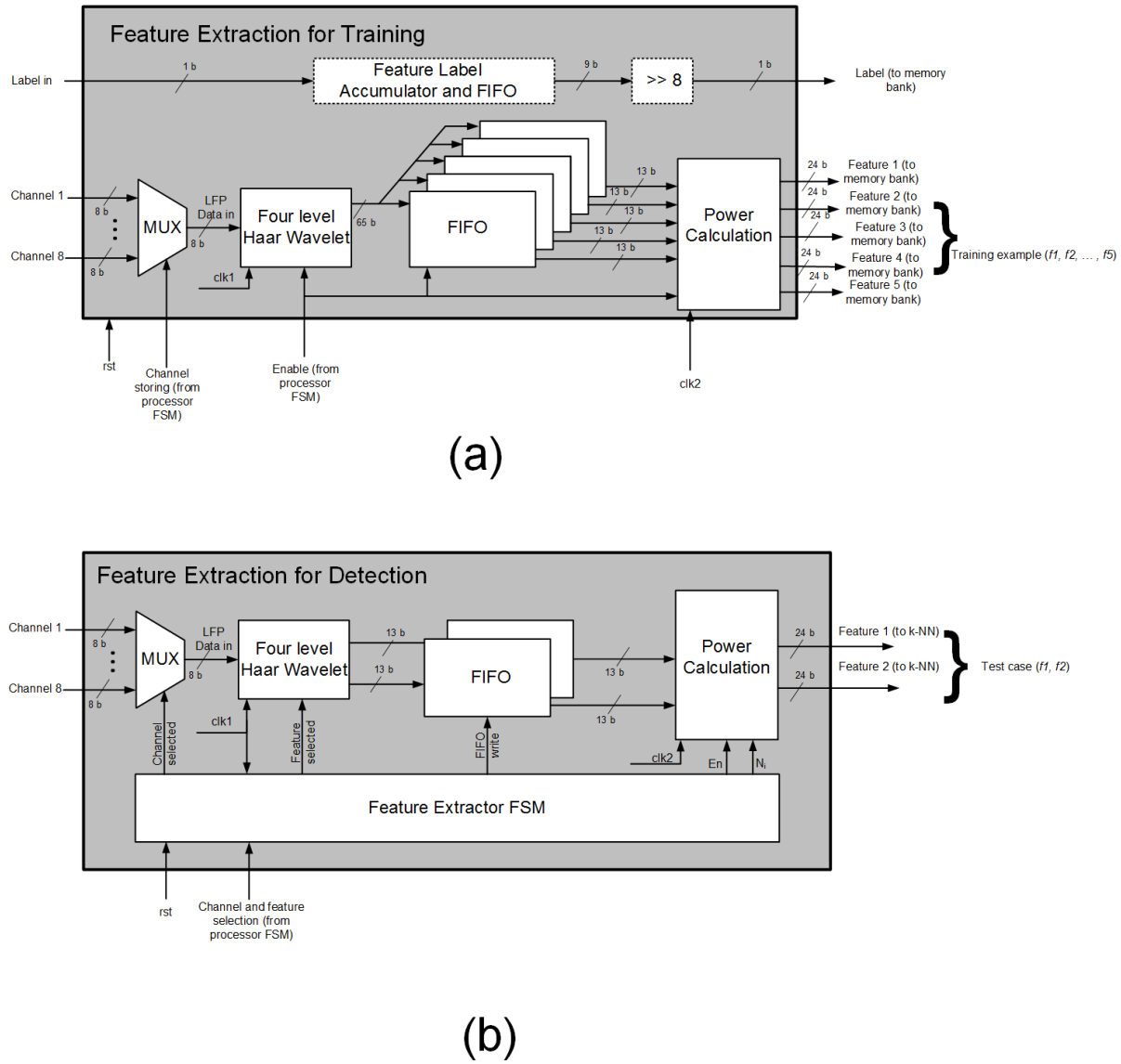


Figure 5.4: Feature extraction units. (a) Feature extraction for training. (b) Feature extraction for detection.

'Feature label accumulator and FIFO' block collects ground truth information from training data. Each epoch of training data consists of a stream of labels for input samples which are used as ground truth information for training purposes as well as to validate performance.

Feature Extraction for Detection

The features are obtained using a 2-second window of LFP signals with 50% overlapping windows. Since, 50% of buffered coefficients are reused the feature extractor FSM controls reading samples from the FIFO. However, the number of samples buffered varies depending on the features selected. The feature extractor controls the reading and writing based on the feature selected. Channel and feature selection is communicated to the feature extractor FSM from the PD processor FSM. In addition, the 'Feature extractor FSM' controls power calculation. The computation of average power of the coefficients in each level varies depending on the feature selected, because the number of samples N_i used in computing the average reduces by a multiple of two for every decomposition level. The feature extractor FSM provides the N_i, which is the number of samples required for computing the average power. The obtained features are sent to the classifier for classification.

5.2.1.2 Four-Level Haar Wavelet

The decomposition represents filtering using binary-scaled frequency bands that maintain the temporal information present in original signal. Figure 5.5 (a) shows the configuration for obtaining the approximation and detail coefficients for a four-level DWT decomposition. To obtain the detail and approximation coefficients at the i-th level, approximation coefficients from the previous level a_{i-1} serve as input as is shown in Figure 5.5 (a). These are convolved with a half-band low pass filter (LPF) h_0 generating the approximation coefficients a_i, and with a half-band high-pass filter (HPF) g_0 to generate the detail coefficients d_i. The convolution and downsampling operation at each level can be represented mathematically as,

$$a_i(k) = \sum_n h_0(n) \cdot a_{i-1}(2k - n) \tag{5.2}$$

$$d_i(k) = \sum_n g_0(n) \cdot a_{i-1}(2k - n) \tag{5.3}$$

where n represents the index of the filter coefficients (low and high pass filters), k is the index of the input signal (approximation coefficients, a_{i-1}). The first part of the equations for both approximation and detail coefficients show a down-sampling by two at each level before filtering. This was to discard redundant information that must have been acquired in the previous decomposition level. Computationally efficient Haar wavelet adopts the polyphase implementation [211]. This is shown in Figure 5.5 (b). The filters for the Haar wavelet are simple 2-tap FIR filters given by $g_0 = \frac{1}{\sqrt{2}} \cdot [1, -1]$ and $h_0 = \frac{1}{\sqrt{2}} \cdot [1, 1]$. The symmetry of the coefficients is exploited by the polyphase implementation in Figure 5.5 (b) such that a single multiplier is shared between both

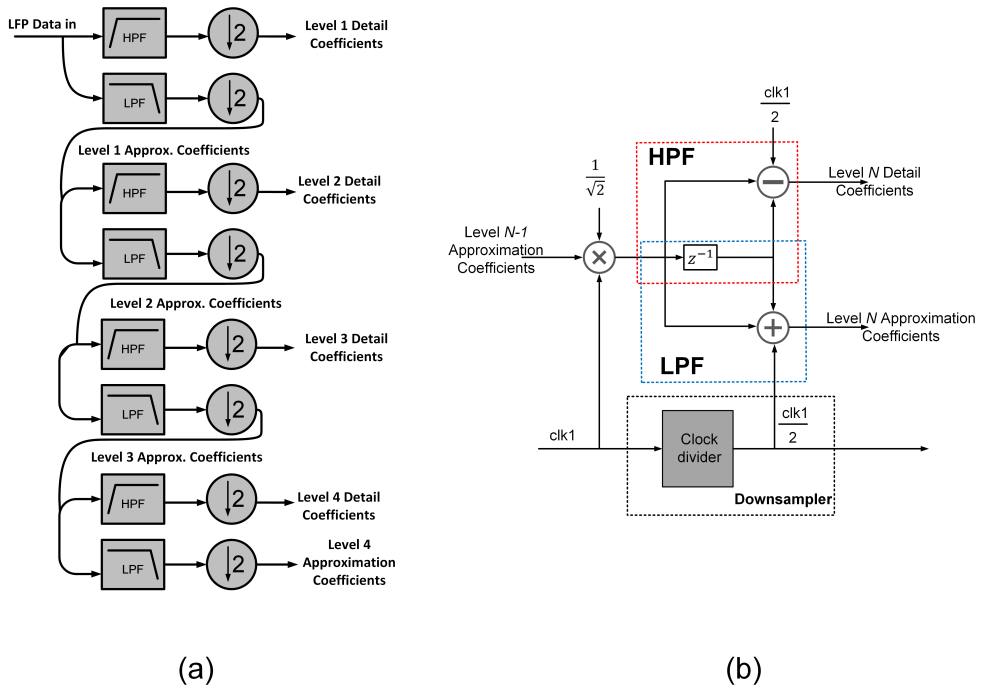


Figure 5.5: (a) Four-level Haar wavelet decomposition. (b) Hardware implementation of Haar wavelet for each level.

filters. The frequency of the input clock *clk1* is divided by two to generate a new clock that is used in controlling the adder and subtractor in the LPF and HPF block respectively. The newly generated clock is also used as the input clock to the next level of the four-level Haar-wavelet. The down-sampling at each level implies that half of the outputs are discarded at each level. This means half of the computations can be avoided. The implementation process exploits this in computing the DWT coefficients of the HPF as well as that of the LPF. It starts by splitting the data stream into odd and even streams. Filtering is implemented such that only the outputs that need to be retained are computed. This approach is depicted in Figure 5.5 (b). The lower part implements the low pass filter and the upper part implements the high pass filter. The adder and delay toggle at every clock cycle. The polyphase implementation ensures that computation occurs half of the time by taking advantage of the down-sampler.

5.2.2 Memory Bank

To facilitate concurrent training and detection, and proper sharing of other computational resources, the PD detection processor consists of a memory bank that is used for storing training data. The major functional units of the memory bank are shown in Figure 5.6. The memory bank unit includes eight channel random access memory (RAM), and static RAM (SRAM) write control and an SRAM read control. Each channel RAM is used for storing training data from its corresponding channel. The memory locations of the channel RAMs are scalable depending on the number of training examples required. The number of training example stored depends on the implementation of the PD detection processor (mainly determined by the k-NN configuration adopted). The training data for each channel is ensured to be made up of half PD examples and

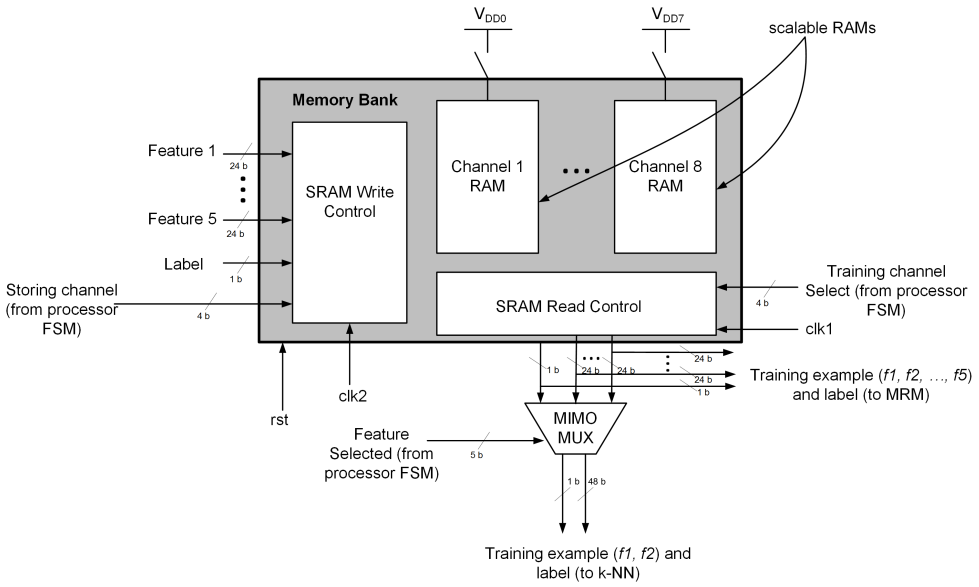


Figure 5.6: Functional units of the memory bank.

the other half consisting of non-PD examples. This is to ensure proper generalizability for both PD and non-PD events. Training data storage is controlled through the SRAM write control. The input data width for each memory is 121 bits (five features of 24-bits each with a corresponding 1-bit label as ground truth). Data writing is controlled through the 'storing channel' control signal from the processor FSM. The data write frequency *clk2* is 1 Hz. The input data consists of five features (24 bits each) and a label (1-bit). Data reading is also controlled by the processor FSM through the 'Training channel select' control signal and data is read at a frequency *clk1* = 128 Hz. Data writing uses a slower frequency because it occurs in real-time, while reading data from the memory bank occurs during MRM and k-NN training.

For MRM training, 121 bits data (consisting of five features and a label are used for training). While the k-NN training data depends on the feature and channel selected after MRM training which is controlled by the processor FSM. Training data for k-NN is obtained through the multiple input multiple output (MIMO) multiplexer, and for each training example 49-bits (two 24-bit features and a 1-bit label) are sent to the k-NN unit for training. The memory bank adopts power and clock gating to reduce leakage and dynamic power respectively when the units are not in use.

5.2.3 Feature Selection

As stated in the Section 5.2.1, feature extraction obtains the most useful information for classification. However, in hardware-efficient applications, an additional stage is required to reduce the number of features used for PD detection. This reduces computation complexity. Feature selection chooses the most relevant features so as to reduce the memory and computational resources. Also, it reduces data over-fitting at the classification stage, since some features are noisy and can lead to degradation in classifier performance. Feature selection results in significant reductions in the area and power of the PD detection hardware. For multichannel application, the classification

phase has to be trained channel by channel either in sequence or parallel. And the best performing channel may be selected for use. This can be computationally intensive. Alternatively, the MRM estimates the most informative channel and uses the features from this channel; as against using all channels [90].

Figure 5.7 shows the function units of the MRM: accumulator bank, MRM division and feature sorting, channel sorting and selection, and the MRM FSM. The 'MRM FSM' controls the MRM training for each channel and the ranking of channels based on the separability of their features by specifying the number of features required to measure separability between PD and non-PD events for each channel. This is done using the 'number of features' control signal from the MRM. Also, feature and channel ranking is done in a time-multiplexed manner to enable logic reuse. The MRM process starts by loading the training data of each channel sequentially. For each channel, training examples for PD and non-PD events are accumulated and stored in the PD and non-PD coefficient stack in Figure 5.7. The MRM dividend and divisor are determined based on the MRM process described in Chapter 4, for each of the five features of PD and non-PD training examples. The 'division and sorting FSM' controls this. Both PD and non-PD events have five features. The separability of each feature is determined using the ratio of PD and non-PD examples for that feature. This is done sequentially for each of the five features using the 'efficient division of unsigned numbers' as demonstrated in [245]. After the ratio of all the features for each channel is obtained the channel weight is obtained using the 'feature sorting and channel weighting' unit. The Channel weights and feature ranks are sent to the 'channel sorting and selection' unit, where the channel weights and feature ranks of each channel are stored. The process is repeated for the other channels until the channel weight of all the channels are obtained. After which it is used to assess the most separable channel and features for use in PD detection. This is communicated to the processor FSM through the 'MRM selected channel' and 'MRM selected feature' signals in Figure 5.7. During the entire process of MRM, the MRM stage is continuously communicating to the processor FSM through the 'MRM state' signal. The MRM feature selection is computationally simple and estimates the most separable features and channel. The only intensive operations it implements are the bubble sort, and efficient division of unsigned numbers. To rank features and channel weights, the MRM requires a sorting algorithm. Efficient sorting is required to limit the time as well as computational complexity required. The bubble sort algorithm was adopted. It sorts data by repeatedly stepping through the list to be sorted and comparing adjacent samples and swapping them if they are in the wrong order. The process is repeated until no more swaps are required. It was adopted throughout the PD detection processor for sorting because it requires a maximum of n^2 clock cycles to complete the sorting task (where n is the number of samples in the list). In addition, it only requires one extra memory location (temporary register) to enable data swap. This makes it time and resource efficient for sorting a small number of data set. More details on other sorting algorithms are provided in [246]. Efficient division of unsigned numbers in [245], was adopted as the power features consisted wholly of unsigned numbers.

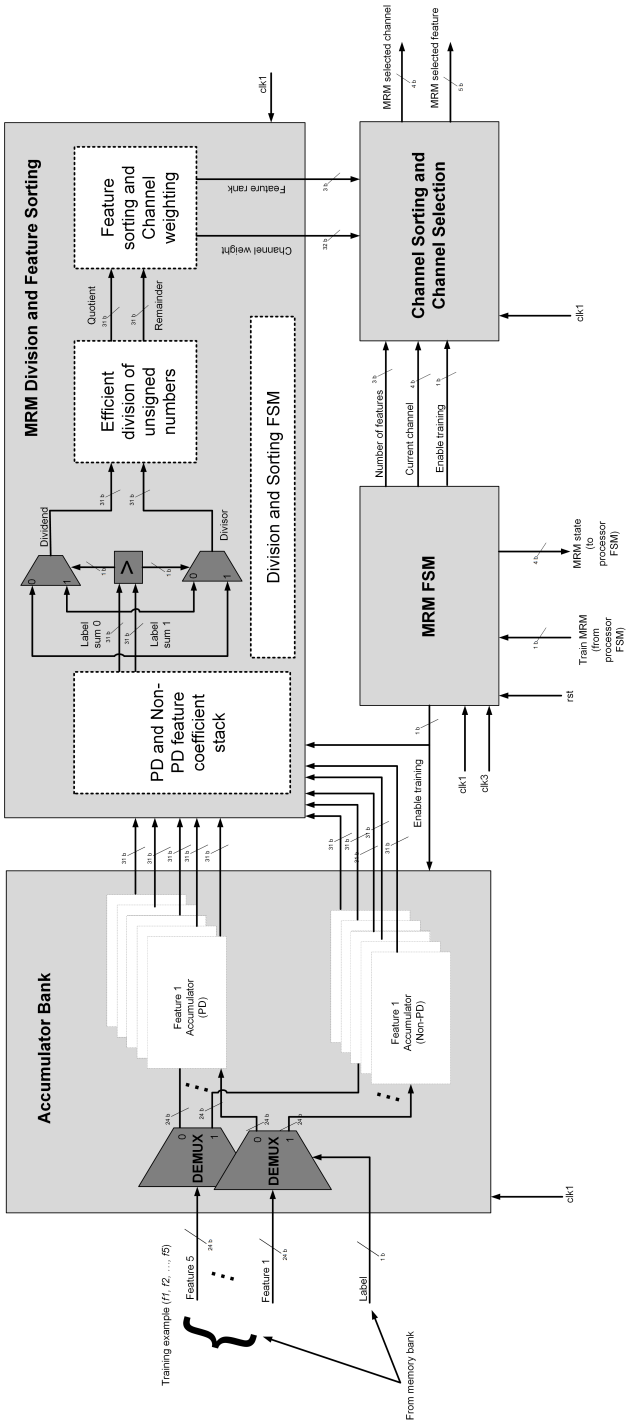


Figure 5.7: Functional units of the MRM.

5.2.4 PD Classifier

Machine-learning algorithms offer the benefit of understanding disease progression in patients. However, their computations are not well supported by conventional DSP platforms; particularly when high order models are used. That is why pre-processing stages like feature extraction and feature selection are used to reduce the computational demands of detection algorithms. In PD monitoring, the classifier is required to identify PD events from non-PD events using acquired neural signals. The PD classifier produces binary outputs, with one representing a PD event and zero representing a non-PD event. The detection results are intended for use either by caregivers or interfaced to stimulation devices. PD monitoring and event detection can be implemented either by using online or offline classification. In offline classification, obtained features or neural samples are transmitted to external devices for classification. In other cases, features are processed internally but other parts of the processing at the classification stage is delegated to an external device. As a result, offline classification can be partly offline or completely offline. For partly (hybrid) offline classification, the training phase (which is computationally intensive) is implemented externally from the system due to its computational complexity; with detection implemented in real-time. On the other hand, in completely offline classification, the classifier is external to the system. Offline classification employs powerful classification algorithms. This makes them more accurate than online classification algorithms. However, for real-time implementation, transmission of raw data imposes power and bandwidth constraints which may be undesirable for fully implantable systems. Also, their computational and memory storage requirements make them impractical for fully implantable bio-electronic systems. Online implementation of PD classifiers allows real-time operation but at the cost of a reduced accuracy due to the power and area constraints of online processing. To facilitate the shift towards fully implantable on-demand DBS systems that can be used to monitor and mitigate PD in real-time, this work deployed an online classifier for PD detection. The PD classifier adopted a k-NN classifier based on the evaluation conducted in Chapter 4. The overall architecture of the k-NN classifier is shown in Figure 5.8. The k-NN classifier consists of four functional units: training examples and label block, level median stack, k-NN controller and kd-tree block. Training data from the memory bank is stored in the training examples and labels block. This data is used to obtain the various medians for different levels of the kd-tree search (nearest neighbours are obtained using kd-tree search).

The medians in the level median stack are used for kd-tree implementation of the k-NN classifier. The k-NN classifier uses a kd-tree approach. The kd-tree was chosen because it uses a hardware efficient implementation of k-NN. The k-d tree distance metric uses the minimum number of computation even though it comes with a reduced accuracy. It is for this reason this platform investigates its use in various k-NN configurations. The performance of various implementations of kd-tree are investigated using different levels of the kd-tree search as well as nearest neighbours. The kd-tree implementation in Figure 5.8 is that of a complementary k-NN which uses majority voting between

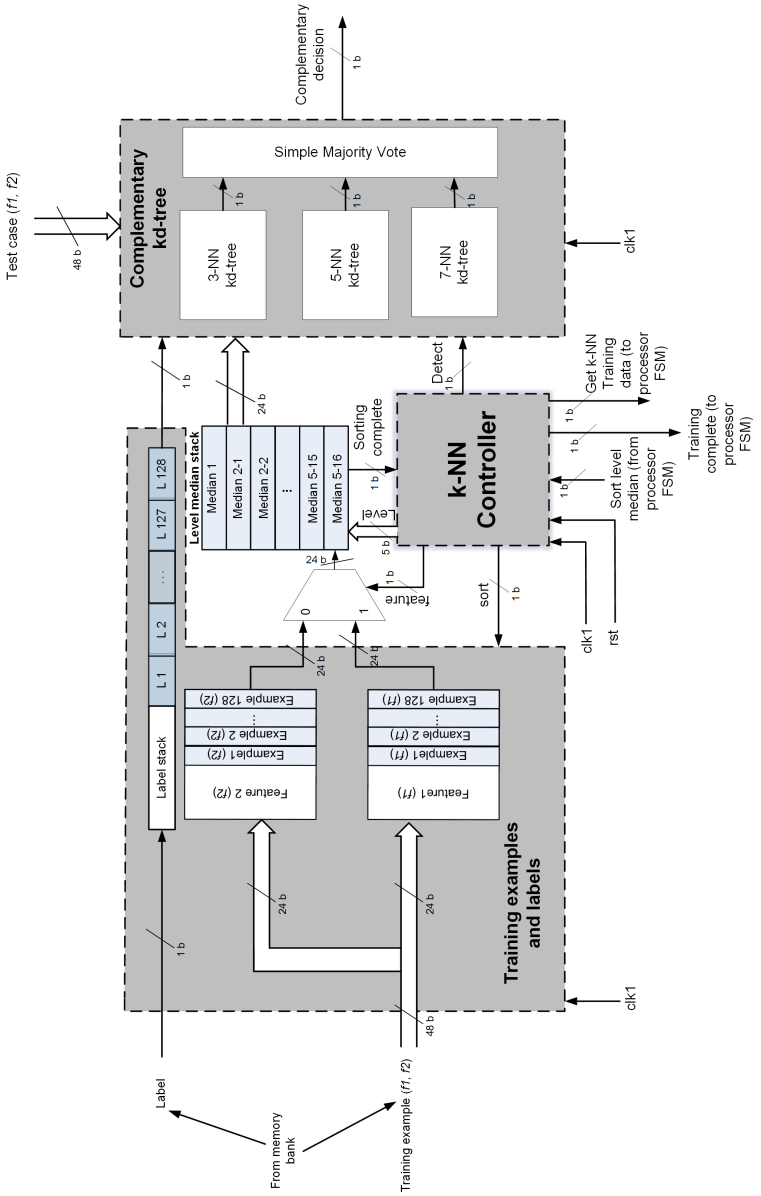


Figure 5.8: Functional units of the complementary k-NN block.

different implementations to produce a classifier with a strong confidence level from three disparate and weak classifiers. This is particularly important since the kd-tree implementation uses an approximate distance measure, as such the complementary configuration compensates for this. The following section provides more detail on kd-tree implementation.

5.2.4.1 k-dimensional Tree Implementation

PD classification using k-NN usually compares the input feature vector to the k-closest training examples. This distance is computed using different metrics, namely; Euclidean [247], L^P distances [248], Mahalonobis distance [249], and approximate distance metrics like k-d tree implementations [250]. Among those mentioned above, the k-d tree implementation is the least computationally intensive. The k-d tree uses a binary decision tree to drill down n-levels, with each level generating a splitting hyperplane that divides the space into two parts, known as half-spaces. The splitting hyperplane is chosen such that every node in the tree is related to one of the k-dimensions (in our case two-dimensions, since we have two dimensional features) and its direction is perpendicular to the axis it splits as indicated in Figure 5.9 (a). Figure 5.9 (a) shows a four-level kd-tree search. The two dimensional features in Figure 5.9 (a) are level 1 detail power (32 49.5 band) and level 2 detail power characteristics (16 32 band). The feature space is divided into 16 compartments by the kd-tree search. For N level k-d tree, there are 2^N nodes, and as shown in Figure 5.9 (b) each test case has to visit N nodes before it is classified. In Figure 5.9 (a), for level 1, a splitting hyperplane is chosen at the median of the x-axis (feature 1) values, all points with a value smaller than the median are categorised in the left sub-tree (consisting compartments labelled I - P) and those greater than the median are categorised to the right tree (with compartments labelled A - H). In this situation, the hyperplane is determined by the x-axis. However, for level 2 splitting, two new hyperplanes are used to further split the left and right compartments of the feature space into four compartments. The process of binary splitting returns to the x-axis in level-3, further splitting each of the four compartments into eight compartments. Level 4 splitting then splits the eight compartments into 16 compartments. These 16 compartments are shown for the feature space of dataset C, with each compartment having four training examples making a total of 64 training examples. Figure 5.9 (a) shows how the 3-NN can be used to classify a test case, as the three closest training examples from compartment 'H' are used. However, to use 5-NN or 7-NN classification, only three level kd-tree can be used. As such, for 5-NN or 7-NN classification, the nearest neighbours from compartment 'G' and 'H' in Figure 5.9 (a) are used in classifying the test data. In the k-d tree implementation, each node is a hyperplane, and at each node, a decision is taken to go either left or right on the tree as can be seen in Figure 5.9 (b). The two feature test data starts by testing feature 1 (f1) at the first level to determine where to go, for the kd-tree in Figure 5.9 (b), if f1 is less than the level-1 median (M1), it takes the 'yes' route otherwise it takes the 'no' route. The process is continued by alternating between features $f1$ and $f2$ at each level until it reaches a node which is at level-4. The various nodes in the kd-tree are shown in Figure

5.9 (b) (A - P from right to left). Their respective compartments are shown in Figure 5.9 (a) on the feature space. This work uses a maximum of five level kd-tree. Various implementations for the 3-NN, 5-NN, 7-NN and the complementary k-NN (which will be indicated by X-NN from now onwards) are discussed in the following section.

5.2.4.2 k-NN Implementations

For each input test case, classification begins hierarchically from the root node in Figure 5.9 (b) and terminates at a leaf node, where a label is assigned (either PD or non-PD event). Different levels of kd-tree search, as well as nearest neighbours, are implemented and tested for accuracy and hardware resources trade-offs. The classifiers studied include k-NN with three, five and seven neighbours; and an ensemble of the classifiers which is tagged complementary configuration. The 'k-NN controller' block in Figure 5.8 controls the sorting of features in the training examples and the storing of the hyperplane (median) points for each level of the k-d tree. In the complementary configuration, a multi-classifier vote is adopted based on the outputs from each of the three order classifiers and the simple majority vote is used to generate a consensus. Below is a brief description of the various implementations.

1. *3-NN kd-tree*: The 3-NN kd-tree implementations are designed to terminate the search at nodes with three training examples such that each test case is classified according to the class dominated by the node it falls into using simple majority vote. Implementations using three, four and five levels kd-tree using 24, 48 and 96 training examples respectively, are investigated for accuracy and hardware resource trade-offs.

2. *5-NN kd-tree*: The 5-NN kd-tree implementations are designed to terminate the search at nodes with five training examples such that each test case is classified according to the class dominated by the node it falls into using simple majority vote. Implementations using three, four and five levels kd-tree using 40, 80 and 120 training examples respectively, are investigated for accuracy and hardware resource trade-offs.

3. *7-NN kd-tree*: The 7-NN kd-tree implementations are designed to terminate the search at nodes with seven training examples such that each test case is classified according to the class dominated by the node it falls into using simple majority vote. Implementations using three, four and five levels kd-tree using 56, 112 and 224 training examples respectively, are investigated for accuracy and hardware resource trade-offs.

4. *Complementary kd-tree*: In the complementary configuration, a multi-classifier vote is adopted based on the outputs from each of the three order classifiers (3-NN, 5-NN and 7-NN). A simple majority vote is used to generate a consensus between the three classifiers. The kd-tree implementation of this consists of nodes with four training examples such that 3-NN could be obtained from the final nodes, while 5-NN and 7-NN are obtained from the last but one level, since it is

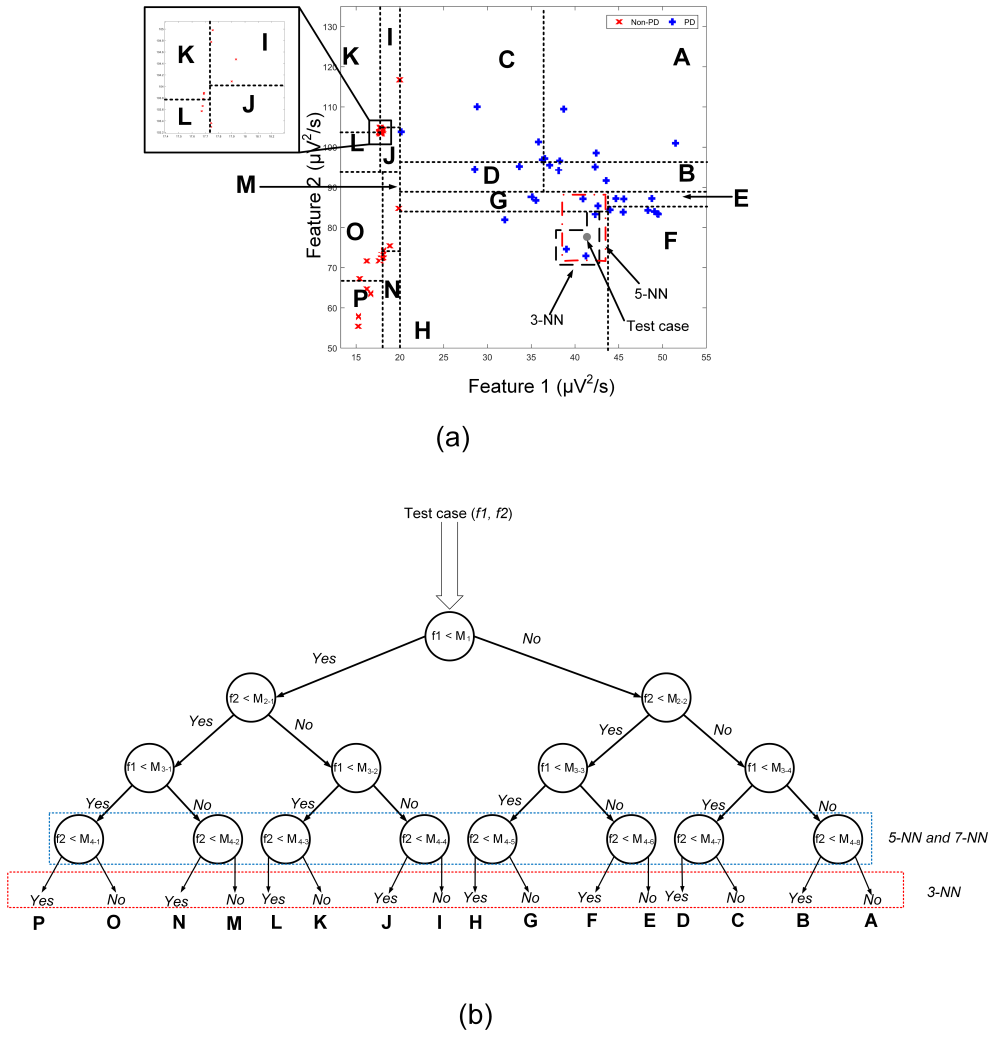


Figure 5.9: (a) Feature space depicting kd-tree compartmentalisation. (b) kd-tree search method.

a node with eight training examples. A typical example of this implementation is in Figure 5.9 (b), which uses 3/4 kd-tree levels. Whereby 5-NN and 7-NN are obtained at level 3, while 3-NN is obtained at level-4. Implementations using 3/4 kd-tree levels have 64 training examples and implementations using 4/5 kd-tree levels have 128 training examples are investigated for accuracy and hardware resource trade-offs. Table 5.1 summarises the various kd-tree implementations and their training data requirements.

5.2.5 Operating Modes

The PD detector has two modes of operation: concurrent detection and training, and detection only. During both modes, real-time PD detection continues and both operating modes use fully on-site computation. During concurrent detection and training, the PD detector concurrently detects PD events and trains to determine the appropriate input channel and features for PD monitoring as well as the k-d tree hyperplanes for each level. This mode requires the most computing and memory requirement. Never-

theless, it only lasts for 5.5% of the time. It involves three sub-modes: training example storage, MRM training and k-NN training. These sub-modes take 4%, 1% and 0.5% of the time respectively. Table 5.2 summarises the active functional units during each of the sub-modes of concurrent detection and training. The second mode is PD detection only, which happens most of the time. During this mode, the PD detector transmits PD events at one-second intervals. Only three functional units are active for majority (94.5%) of the time as summarised in Table 5.2.

Table 5.1: Summarising the various kd-tree implementations investigated.

k-NN Implementation	Kd-tree Levels	Number of Training Examples	Relative Training Frequency *
3-NN	3	24	5.33
	4	48	2.67
	5	96	1.33
5-NN	3	40	3.20
	4	80	1.60
	5	160	0.80
7-NN	3	56	2.29
	4	112	1.14
	5	224	0.57
X-NN	3/4	64	2
	4/5	128	1

* Training frequency measured relative to X-NN (4/5), which is trained once a day.

Table 5.2: Operating modes of the PD detection processor.

Operating Mode	Sub-mode	Period ($\times\frac{1}{clk3}$) **	Active Functional Units (aside from processor FSM)
Concurrent detection and training	Training example storage	8	Feature extraction for training, memory bank, feature extraction for detection and k-NN.
	MRM training	2	Memory bank, MRM, feature extraction for detection and k-NN.
	k-NN training	1	Memory bank, feature extraction for detection and k-NN.
Detection only	–	189	Feature extraction for detection and k-NN.

** $clk3$ varies for each implementation. For X-NN (4/5) it is $\frac{1}{432}$ Hz: this frequency has to be multiplied by the relative training frequency in Table 5.1 to obtain the equivalent value of $clk4$ for other implementations of the PD detection processor.

5.3 Physical Implementation

The objective is to provide a hardware platform for real-time processing of acquired neural data such that PD events can be distinguished from non-PD events. A fully

online implementation performs on-site and real-time PD detection such that only PD events are transmitted to caregivers or stimulation devices to trigger actionable outputs. The most suitable candidate is an FPGA-based platform since it provides more flexibility for investigating various implementations of the PD detector as against an ASIC implementation which is hardwired. Currently, the mechanisms of PD are still under debate and the PD detection algorithms may need to be updated as a deeper understanding of the mechanisms of PD are gained. In essence, an FPGA based platform offers the best compromise between adaptability and portability. Moreover, register transfer level (RTL) implementations used on FPGAs are easy to translate for ASICs. The PD detection processor was modelled using Verilog hardware description language (HDL), and was programmed on an Artix-7 FPGA as shown in Figure 5.10 (a). The FPGA based PD detection tool was implemented in three major layers: the MATLAB layer, universal asynchronous receiver/transmitter (UART) layer and the FPGA layer. This is shown in Figure 5.10 (a). Synthesised LFP data is transferred from MATLAB to the FPGA board through the serial communication link (USB-JTAG cable in Figure 5.10(b)) to the Artix-7 FPGA for processing. Processed data is sent back to MATLAB for performance evaluation of the PD detector (using signal ground truth stored in MATLAB). For serial communication, the UART layer interfaces MATLAB and FPGA, using a UART script in MATLAB and an interfacing UART receiver and transmitter implementations for deserialising and serialising input and output data from the FPGA respectively. The UART layer mainly controls data flow between MATLAB and the FPGA platform as is shown in Figure 5.10 (a). The following section further provides details on the justification for FPGA implementation, the target hardware selected (FPGA board) for validating the platform and the input-output interface (UART) adopted.

5.3.1 FPGA Implementation

FPGA is hardware platform that is configured using HDL. Compared to micro-controllers and ASIC, FPGA provides the best trade-off between speed and flexibility. ASICs are known to have a fixed architecture, since each chip is designed strictly to perform a specific function. They are designed such that power and area consumption are optimised. This limits their flexibility but improves their speed and possibly their efficiency. On the other hand, micro-controllers are highly flexible because they can be programmed and reprogrammed to perform almost any operation. However, their inherent and singularly defined structure for performing programmed instructions limits their operating speed as every programmed instruction is performed in a nearly similar pattern [245]. FPGAs have a blend of the fixed architectures in ASIC and the structured programme execution style synonymous with micro-controllers; this makes them in between the two extremes regarding speed and flexibility and serves as a good compromise. They are manufactured such that they contain standard hardware building blocks like: look-up tables (LUT), memory, adders, multipliers and a reconfigurable interconnect structure that enables connection between its different blocks. Their ability to be programmed using HDLs to perform complex sequential and combinational logic functions in a sim-

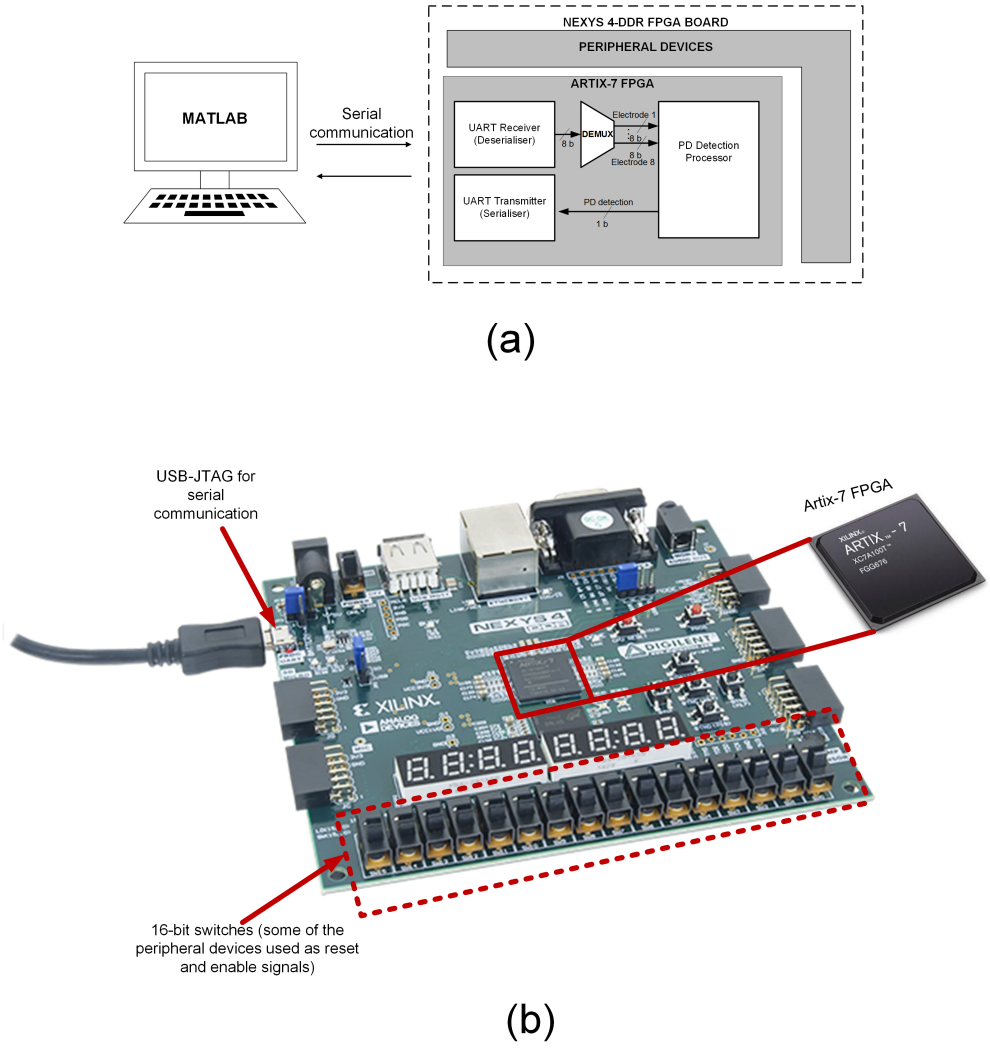


Figure 5.10: FPGA implementation of PD detection.

ilar fashion as ASICs results in their speed and flexibility compromise. It is for this reason that the hardware-efficient implementation of the PD detector is validated on an FPGA platform. Also, since it is structured to perform complex operations in parallel like in ASIC and the long term goal is to implement fully online and real-time implantable ASICs that can be deployed for PD monitoring and DBS modulation, it serves as a 'semi-ASIC' platform. Among all the reasons that influenced the choice of an FPGA based PD detector, of paramount importance is the need to implement and validate on a platform that provided room for modifying the PD detection algorithm such that it can be updated and performance of various functional units investigated.

5.3.2 Targeted Hardware

The targeted hardware is the Nexys4 DDR processing board from Xilinx with Artix -7 FPGA. The Nexys 4 DDR uses its own expansion system and has 60 I/O pins that can be interfaced to external devices as is shown in Figure 5.10 (b). The board uses 3.3V I/O. It has on board peripheral devices that are accessible to the Artix-7 chip as peripheral I/O devices. Its high capacity FPGA is the XC7A100T-1CSG324C. Table

5.3 summarises the FPGA packaging and device details based on the guide from [251]. Apart from the FPGA chip, the Nexys4 DDR boards consists of a number of peripherals devices such as, LEDs, switches, temperature sensor, accelerometer, a speaker amplifier, microelectromechanical systems (MEMS) digital microphone, and a number of input and output devices that enables it to be used with a host of interfaces [252]. Of primary importance is its USB-JTAG port which was used for FPGA programming and data streaming through UART as is shown in Figure 5.10 (b). For communication with a host PC and programming the Artix-7 FPGA, it uses a USB mini-B connector. Its configuration bit files are held in a 128Mbit serial peripheral interface flash chip. Finally, it has a 128 Mbit synchronous dynamic RAM chip on board.

Table 5.3: FPGA Packaging and device details.

FPGA information	Device type	Speed grade	Package type	Lead-free package	Number of pins	Temperature range
Code	XC7A100T	-1	CS	G	324	C (commercial, between 0^0C to 85^0C)

5.3.3 Input/Output Interface

The synthesised LFP signals used as input data are described in Chapter 4, and are fed from MATLAB. The LFP data is initially represented in double precision floating point number. It is then quantized to 8 bits fixed point representation, with one sign bit and one fractional bit. Serial communication is implemented using UART protocol, in a full-duplex configuration. From Figure 5.10 (a), the input data from MATLAB is divided into packets of 8 bits which are sent in serial format to the FPGA platform. The UART receiver block in the FPGA receives the input serial data and deserializes it. The deserializer block implemented in Verilog HDL takes in serial input data from MATLAB, buffers them until a complete word is obtained before it sends it as input to the PD detection processor. Also, the UART data packets are sent with a channel identifier packet to determine the recording channel from which the LFP signal originates from. The obtained packet and channel identifier are used to demultiplex the input signal to the appropriate channel of the PD detection processor for PD detection. After PD detection, labels are assigned to input signals. The binary labels are sent through the UART transmitter block of the FPGA back to MATLAB as PD and non-PD events. The complete architecture of the data streaming interface is shown in Figure 5.10 (a).

5.4 Results

To evaluate the performance of the proposed PD detection method, semi-synthetic neural signals constructed from the real LFP recordings as described in Chapter 4. Semi-synthetic neural signals were used as against experimentally recorded neural sig-

nals because all underlying conditions in original LFP signal can be modelled alongside ground truth information. In addition, semi-synthetic signals offer the opportunity to synthesise longer LFP recordings from real LFP recordings with short duration. The results describe the detection performance as well as hardware resource utilisation. Detection performance results were obtained from FPGA while hardware resources were obtained both from FPGA and post-synthesis power and area estimates when mapped to 45 nm CMOS process.

5.4.1 Detection Performance

For hardware test, performance measures were obtained over a complete training and test period $clk4$. It has a period of $\frac{1}{86400}$ Hz for the X-NN (4/5) implementation and for other implementations, the frequencies are multiples of X-NN (4/5) $clk4$ as summarised in the relative training frequency in Table 5.1. Before performance evaluation, training is conducted which lasts 5.5% of the time. Then for the other 94.5% of the time, test cases were detected. The bar plots in Figure 5.11 represents the bar plots of the F1-score for the various implementations. It can be seen that the F1-score increases with an increase in the number of training examples, except in the case of 5-NN(5), 7-NN(3), X-NN(3/4) and X-NN(4/5). In the case of 5-NN(5), this could be attributed to outliers in the training examples due to an increased number of training examples. Outliers tend to bias the classifier model. In the case of 7-NN(3), the use of seven nearest neighbours in a training set consisting of 56 training examples can increase susceptibility to outliers. That is, if outliers make a large part of the training examples.
With a median F1-score of 95.02%, the X-NN(3/4) has a better performance than the 5-NN(5) and 7-NN(4) which have 67.75% and 92.54% respectively. However, they use more training examples than the X-NN(3/4). The improved performance in X-NN(3/4) can be attributed to the complementary detection it uses. For the X-NN(4/5), its median F1-score trails that of the best, which is the 7-NN(5) by less than 1% even though the 7-NN(5) uses 228 training examples compared to 128 training examples used by the X-NN(4/5). The top three implementations in terms of F1-score are: 7-NN (5) with a median of 96.29%, X-NN (4/5) with a median of 95.34% and X-NN(3/4) with a median of 95.02%. The classification accuracy measures in Figure 5.11 (b) follow almost the same pattern as the F1-score. However, the major difference is that the X-NN(4/5) has a better classification accuracy performance than the 7-NN(5). The top three in terms of classification accuracy are: X-NN (4/5) with a median of 91.93%, 7-NN (5) with a median of 89.66% and X-NN(3/4) with a median of 82.14%. From the foregoing, it is clear that the complementary k-NN configuration improves on detection performance. As it outperforms any of the normal k-NN configurations that use about the same number of training examples.

5.4.2 Hardware Resources

Hardware implementation of a PD detector using Haar Wavelet-based FE, MRM based feature selection and k-d tree based implementation of the k-NN classifier were realized using Verilog HDL. This was tested in FPGA. For ASIC resource requirements, various

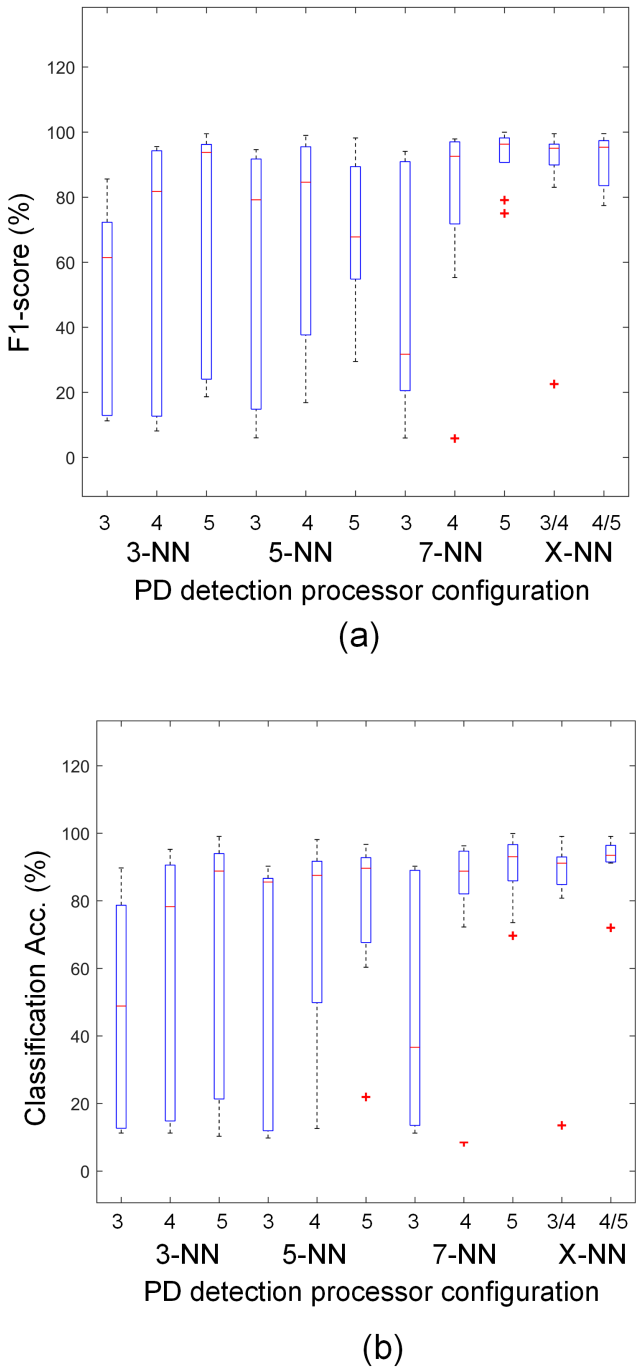


Figure 5.11: Detection accuracy for PD detection processor. (a) F1-score for various configuration of PD detection processor. (b) Classification Accuracy for various configuration of PD detection processor.

implementations of the architecture were synthesised and mapped into a 45nm standard cell library. The following section gives more details on hardware resources.

5.4.2.1 FPGA Resources

The major building blocks for the FPGA implementation are: LUT slices, register slices, multiplexers and DSP units. The Artix -7 FPGA has 63400 LUT slices, 126800 register slices, 240 DSP slices and 47550 multiplexer slices [252]. The resource utilisation for the various implementations are summarised in Table 5.4. It can be seen that only six DSP slices were used. These were used in the feature extraction unit. Other resources increase with an increase in nearest neighbours as well as kd-tree levels. Figure 5.12 (a) and (b) show the proportion of resources consumed by the four major blocks (excluding the PD detection processor FSM). Figure 5.12 (a) shows that of the 3-NN(3) implementation which consumes the least resources, while Figure 5.12 (b) shows that of the 7-NN(5) which consumes the most resources. From Figure 5.12, it can be seen that the k-NN and memory bank dominate in terms of resources as the kd-tree level grows. The feature extraction units and the MRM stay relatively the same throughout. DSP slices are only used in the feature extraction block for multiplication to compute the power of the coefficients from the four-level Haar wavelet.

5.4.2.2 On-Chip Power and Area

The mapped implementation resulted in post-synthesis estimates of dynamic power per channel and area per channel for different configuration of the PD detection processor and are summarised in Table 5.5. On-chip power and area estimates were obtained by mapping the various implementations to 45nm NanGate digital cell library [253]. The power is analysed based on a core voltage of 1.1 V. The 7-NN (3) configuration has the lowest dynamic power requirements, while the 3-NN (3) has the lowest area requirements. Form Table 5.5, it can be seen that the area consumption is commensurate with FPGA resource utilisation. However, this is not the case for the dynamic power consumption, as dynamic power is dependent on the activity in the circuit mostly resulting from control and clock signals.

Table 5.4: Table summarising FPGA resources.

Implementation		LUT		Registers		Multiplexers/ Demultiplexers		DSP	
		slices	(%)	slices	(%)	slices	(%)	slices	(%)
3-NN	3	8377	13.21	6139	4.84	236	236	6	2.5
	4	11071	17.46	7492	5.91	619	1.30	6	2.5
	5	17135	27.03	10140	8.00	1078	2.27	6	2.5
5-NN	3	10284	16.22	6953	5.48	331	0.7	6	2.5
	4	15351	24.21	9033	7.12	987	2.07	6	2.5
	5	23855	37.63	13017	10.27	2512	5.28	6	2.5
7-NN	3	11514	18.16	7656	6.04	696	1.46	6	2.5
	4	17675	27.88	10646	8.25	1674	3.52	6	2.5
	5	29839	47.06	15997	12.62	3636	7.66	6	2.5
X-NN	3/4	14194	22.39	8284	6.53	885	1.86	6	2.5
	4/5	20993	33.11	11560	9.12	1904	4	6	2.5

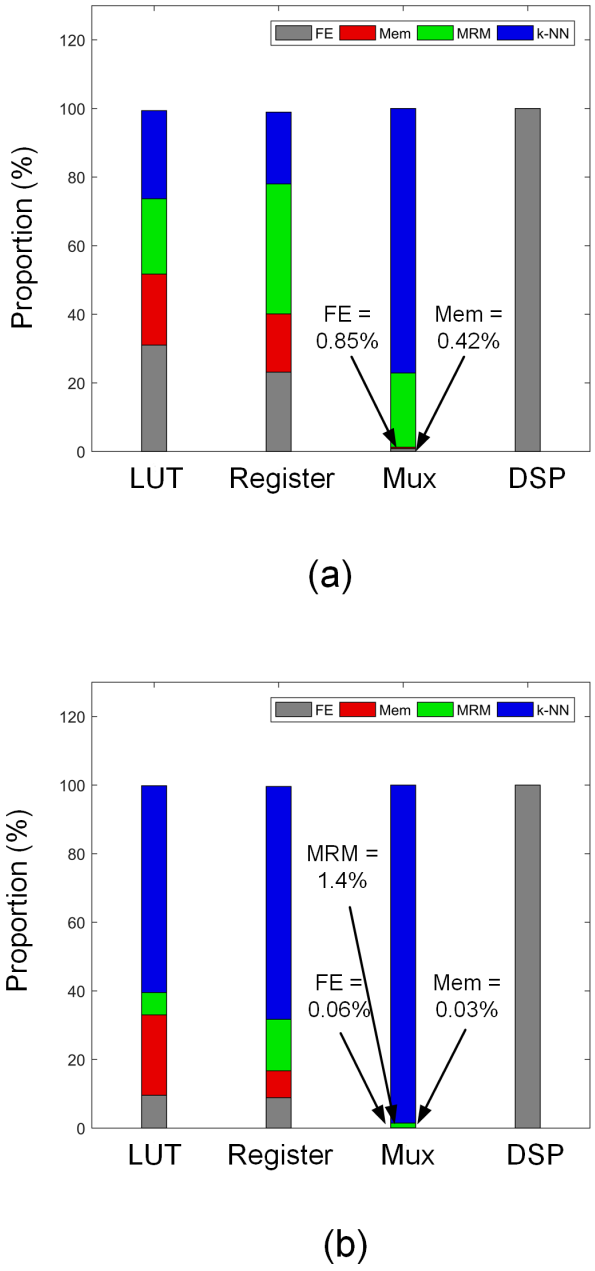


Figure 5.12: FPGA resource utilisation.

Table 5.5: Table summarising performance for various implementations.

S/No.	Implementation		Complexity Measures		Accuracy Measures		
			Dynamic Power ($\mu W/ch$)	Area (mm^2/ch)	Mean MCC	Mean F1 (%)	Mean Classification Acc. (%)
1	3-NN	3	2.28	0.0433	0.0012	51.29	49.74
2		4	2.37	0.1175	0.2479	56.69	55.70
3		5	2.31	0.1915	0.3513	68.20	64.76
4	5-NN	3	2.23	0.0465	0.1608	56.12	60.00
5		4	2.34	0.1651	0.3275	68.43	69.48
6		5	2.53	0.2873	0.5081	68.11	77.96
7	7-NN	3	1.80	0.0496	0.1315	50.20	51.18
8		4	2.49	0.2126	0.482	78.37	79.79
9		5	2.73	0.3815	0.6422	92.68	89.66
10	X-NN	3/4	2.22	0.1414	0.3542	86.18	82.14
11		4/5	2.26	0.2384	0.6162	91.38	91.93

Nevertheless, there is an increase in clock activity for each implementation with an increase in the number of kd-tree levels adopted. To put the complexity measures in perspective, Figure 5.13 (a) shows the MCC vs dynamic power and Figure 5.13 (b) shows the MCC vs area. The various implementations of the PD processor are represented by the serial numbers in Table 5.5.

The plots are divided into four regions: Q1, Q2, Q3 and Q4. Q1 represents the implementations that have an average MCC greater than 0.5 and also have dynamic power lower than the average dynamic power for all implementations. In Figure 5.13 (a), only X-NN (4/5) implementation makes this category. This is the most desirable outcome in terms of dynamic power, that is, having a classifier to produce a model that accurately classifiers test cases as well as having a dynamic power lower than the class average. The other quadrants in Figure 5.13 are: Q2 which represents an average MCC greater than 0.5 and a dynamic power greater than the average which is 2.32 $\mu W/channel$. This represents a good detector but at a higher power consumption than Q1. Q3 represents the unwanted condition, in which the detector consumes more power than the class average and produces an MCC that is less than 0.5, which signifies weak positive correlation.

The fourth scenario, is Q4 where the implementation results in a weak positive correlation ($MCC < 0.5$) and a power consumption less than the average for all the implementations. This scenario may be more desirable than Q2 depending on what is more important between power consumption and detector performance measure in MCC. The X-NN (3/4) produces the best performance for the implementations in Q4. For the area consumption in Figure 5.13 (b), there is an almost linear increase in area with an increase in MCC. Figure 5.13 (b), divides the quadrants in a similar way to Figure 5.13 (a). However, the average area of the implementations is used to divide the y-axis. None of the implementations fall into Q1. In Q2, the 5-NN(5), 7-NN(5) and X-NN(4/5) fall into Q2 with the X-NN(4/5) presenting a better area trade-off than the other two since it utilises relatively less power. Q3 also represents the undesirable

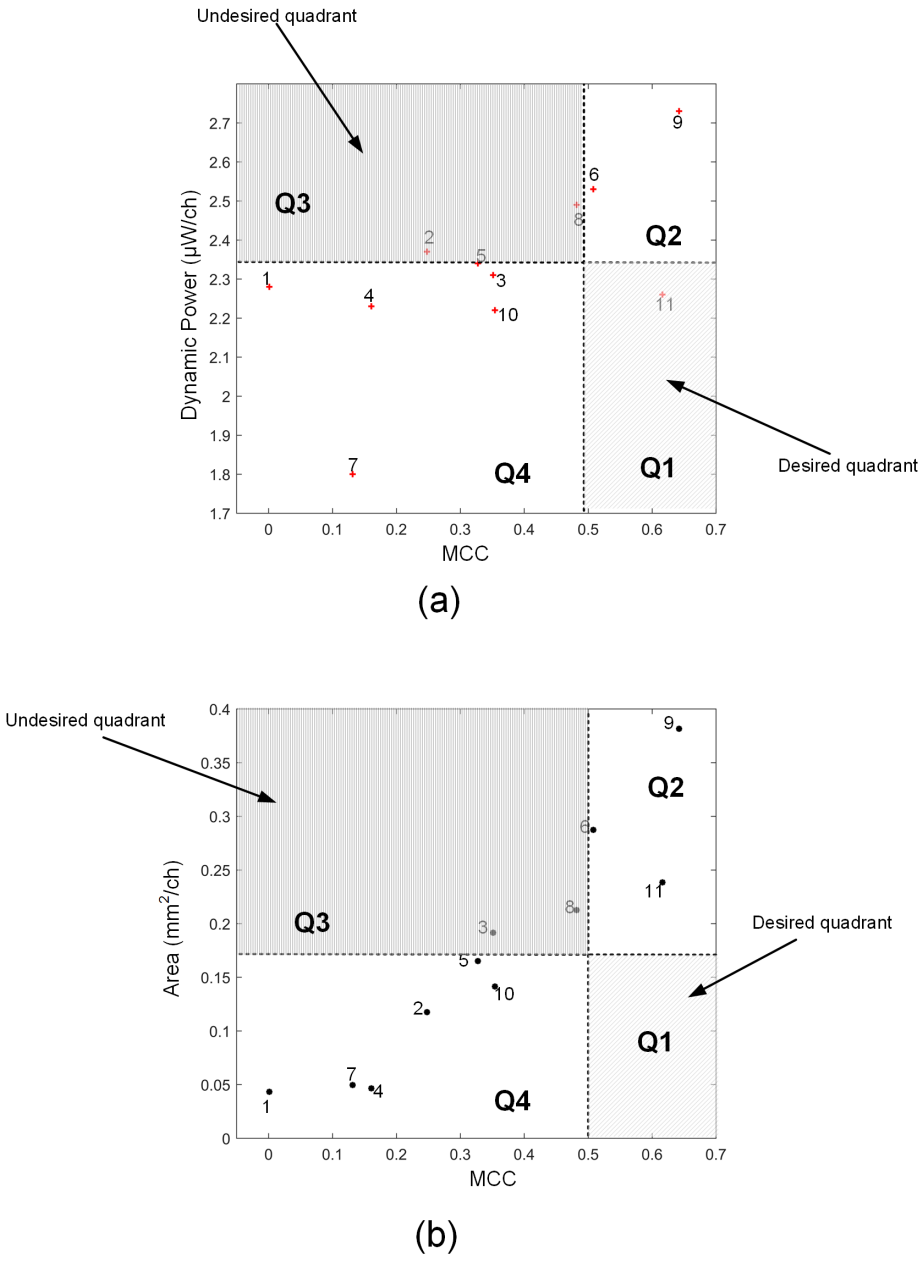


Figure 5.13: On-chip area and power. (a) Dynamic power vs MCC. (b) Area vs MCC.

quadrant, with large area consumption and an average MCC below 0.5. In Q4, the best implementation in terms of MCC is the X-NN(3/4). Thus, this demonstrates that the complementary configuration offers a good trade-off for complexity and accuracy measures.

5.4.3 Comparison with other Neurological Event Detectors

Table 5.6 summarises the characteristics of state of the art neurological event detection processors implemented between 2013 to date. The implementation with the best efficacy, area and dynamic power trade-offs in this work, which uses the X-NN (4/5) classifier is adopted and compared with other neurological event detection processors. It performs on a comparable level in terms of area and power to other neurological event processors as can be seen in Table 5.6. The detection processor developed in this work is the first processor for PD detection. So far, neurological event detection processors have focussed on epileptic seizure detection. For seizure detection, the major signals of interest are EEG and ECoG (mostly < 30 Hz), which are obtained from the scalp and cortex respectively. In seizure detection, most studies use phase synchronisation between signals recorded from different channels as the biomarker of interest. Seizures are detected when phase synchronisation increases above a specified threshold. An increase in EEG or ECoG synchronisation between channels can be indicative of an epileptic episode. The main challenge of using the phase synchronisation approach for detecting epileptic seizures is that the baseband synchronisation varies across subjects and between channels. Nevertheless, neurological signals for epilepsy (EEG and ECoG) are more distinguishable than those of PD (LFP signals). This is why a closed-loop system for epileptic seizure suppression has already gained FDA approval [142]. On the other hand, the unpredictable nature of LFP signals for PD and their higher frequency content compared to EEG and ECoG makes their processing more computationally intensive. LFP signals are used as biomarkers in PD detection and monitoring due to their stability and rich spatiotemporal content [16], which is necessary to overcome the unpredictable nature of PD. Besides, their ability to be obtained from the stimulation electrodes also makes them minimally invasive. However, EEG signals are usually obtained from electrodes attached to the scalp. This makes them impractical for ambulatory disease monitoring as they are not implanted, which means they could limit patients' quality of life. Another issue is that correlation between PD and EEG signals are yet to be established or reported. This made the use of LFP signals necessary.
Fundamentally, the trade-offs are mainly between efficacy and complexity measures. And from Table 5.6 it is clear that in terms of complexity (i.e. on-chip area and power), some neurological events detectors present a better result than the PD detection processor. In terms of on-chip area, the works in [254–256], have a lower area per channel. In terms of power, [254] and [255] present a lower power per channel. Regarding efficacy, most studies in epileptic seizure detection have focussed on sensitivity, which only measures the true-positive rate. This means a randomly guessing processor that always returns a positive can achieve a 100% sensitivity, since it will get all the actual positives (but will achieve a specificity of 0%). This is why more balanced and

comprehensive efficacy measures are required for an objective assessment of processor performance. This is particularly necessary for situations where event detection informs therapy. Administering therapy when it may not be needed can result in side effects [36]. This was the reason more comprehensive measures like the MCC and F1-score are adopted in this work. In addition, the PD detection processor monitors a wider range of frequencies (i.e. < 50 Hz compared to the < 30 Hz in epileptic seizure) which could contribute to its complexity. Nevertheless, it presents superior efficacy than other neurological event detectors. Its complexity is also comparable to those of other neurological event detectors as highlighted in Table 5.6.

Table 5.6: Performance comparison with state of the art neurological event detection processors.

Ref.	This work	[254]	[204]	[257]	[258]	[255]	[256]	[208]	[195]	[259]
Year	2017	2017	2017	2016	2015	2015	2014	2014	2013	2013
Application	PD onset detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection	Epileptic seizure detection
Processing scheme	Four-level Haar wavelet, with feature selection and a complementary k-NN classifier.	Two levels of continuous wavelet transform used to process input signals, seizures detected on threshold crossing.	Seizures detected when phase synchronisation between channels exceeds a certain threshold.	Energy from 7 filter banks classified using a Gaussian basis function SVM classifier.	Energy from 2s overlapping epochs extracted from 15 filter banks and classified using two linear SVM classifiers.	Line length between successive samples detected on threshold crossing.	Energy from 2s epochs extracted through 8 BPFs and classified using SVM.	Entropy and Power spectrum features classified using a linear least square (LLS) classifier	2 second epochs divided into 7 sub-bands (between 0 - 28 Hz). Energy from sub-bands are used as features to an SVM classifier.	FFT features classified using an LLS classifier
CMOS Tech (nm)	45	350	130	180	180	180	130	180	180	180
# of channels	8	32	24	8	16	16	18	8	8	7

Neural Signal	LFP	EEG	EEG and ECoG	EEG	EEG	iEEG	EEG	EEG	EEG	ECoG
Area/ch (mm^2)	0.2384	0.0491	12 [a]	3.125	3.125	0.0625	0.0833	13.47 [a]	25 [a]	0.7436
Core Voltage (V)	1.1	1.25	1.2	1.8	1.8	0.8	1.2	1.8	1.8	1
Dynamic power/ch (μW)	2.26	0.95	1300 [b]	FE = 7, SVM = 12.56	2.45	0.85	5.94	2800 [c]	8.25	4.086
Efficacy (%)	F1 = 91.38, Class acc. = 91.93 and MCC = 0.62 [d]	Sensitivity = 87	Sensitivity = 88 - 96	Sensitivity = 95.1	Sensitivity = 95.7 and Specificity = 98	Sensitivity = 100	Sensitivity = 91 - 96.	Sensitivity = 92	Class. acc = 84.4	Class. acc = 93.1
Configuration	Detection only	Detection only	Includes recording and stimulation	Includes recording	Detection only	Detection only	Detection only	Includes stimulation and communication	Includes recording	Detection only

[a] Gives the total area of the chip not area per channel. [b] total for 24 channels and extra stages. [c] total for 8 channels and extra stages. [d] not in %.

5.5 Discussion

5.5.1 Personalised Health Monitoring in PD

In PD, proper health monitoring can lead to better PD mitigation as well as reduce the number of face-to-face visits by patients. Further understanding on how disease progresses in patient population can be achieved through remote health monitoring. This can involve the analysis of data from multiple patients such that aDBS strategies that incorporate more universal features or biomarkers could be incorporated in tackling varying PD disorders. This will make therapy more proactive and less retroactive. As more and more data is collected a better understanding of disease progression can be obtained, which could be used in refining therapy. Moreover, since diverse population of PD patients are afflicted with various dominant symptoms, implementations of personalised health monitoring systems can be tailored to cater for individual patient needs. Early detection of patient deterioration could help caregivers to immediately take action in modifying stimulation parameters or therapeutic paradigms [260]. This can be achieved by remote and continuous monitoring of the disease states. The information obtained from health monitoring systems are intended to be logged to monitor disease progression or relayed to caregivers in the event of an emergency. This can make therapeutic interventions more sustainable, by early recognition of unresponsive symptoms so that alternative therapeutic regimes are provided [112].

5.5.2 Supporting Technology

Advancements in bio-sensing, signal processing, data communication and nanotechnology are providing the possibility of fully implantable systems for disease monitoring. The major requirement is to devise systems that detect patient-specific physiological states in real-time using minimally invasive and low power devices. This is however made easy because of the development in machine-learning algorithms that are capable of exploiting statistical properties in the data to model specific correlations; which facilitate decision making by healthcare experts and devices [65]. For disease monitoring using machine learning models, prior pre-processing is as important, or even more important than the use of the machine learning models themselves [261]. As the models are only as good as the input signals that are fed to them. This makes the use of high-performance low noise amplifiers (LNA) and analog-to-digital converter (ADC) essential to achieving high-performance detection. In this work, already pre-processed neural signals were used. The PD detection processor was prototyped on an FPGA platform to demonstrate the feasibility of on-site and real-time PD detection without compromising flexibility. Also, translating the RTL implementation to lay-out for on-chip implementation is fairly straightforward. This was why power and area estimates for various implementations of the PD detection processor were obtained for 45 nm CMOS process.

5.5.3 Challenges

Healthcare monitoring systems are required to be operational for long periods of time. Monitoring systems like the LiveNet system [262], which uses an accelerometer, ECG, EMG and skin conductance to obtain information on PD and epileptic seizure, among others, by transmitting raw samples of data to external devices. Adopting this approach imposes high power and throughput requirements. This motivated the adoption of an implementation that can be deployed for on-site real-time processing. Due to the need for detection algorithms to be made implantable, there might be insufficient storage and computing capability to handle large and dynamically changing training data. Large training data is required as classification accuracy improves as more knowledge on the disorder is incorporated into the processor. Other challenges confronted by personalised health monitoring systems are security and privacy issues [263]. So far, the major limitation of disease monitoring is the issue of security and privacy of patient data. Security and privacy is mainly dependent on the adopted communication method used for transmitting PD events. A serious effort is already invested in research on encryption to protect from data interception and tampering [264].

5.6 Chapter Summary

In this chapter, we presented complementary PD detection as a hardware-efficient method for PD detection. Based on the work presented, the following conclusions can be drawn:

a) Complementary detection uses a combination of weak classifiers to produce a classifier with a higher consistency and confidence level than the individual classifiers in the configuration. The complementary configuration was tested against various implementations of the PD detection processor in terms of detection performance, dynamic power and area consumption. The PD detection processor is the first known implementation of a PD detection processor.

b) The design leveraged on the flexibility of FPGA to test the various implementations of the PD detection processor. FPGA implementation offers the opportunity to incorporate additional functionality into the implemented design. In addition, estimates for power and area were obtained for 45 nm CMOS technology.

c) In terms of detection accuracy and optimal power and area trade-off, the complementary detection schemes presented the best trade-offs. Nevertheless, the most optimal classifier configuration is dependent on the emphasis an application gives to power and area in relation to detection performance.

d) The PD detection processor monitors a wider range of frequencies than other neurological event detectors. It presents more comprehensive efficacy measures than other neurological event detectors and it has a complexity that is comparable to state-of-the-art event detectors. This offers a good trade-off in terms of efficacy and complexity.

e) Considering how spectral features are useful as biomarkers in other neurological applications, the proposed approach can find a range of applications and implementations.

f) For hardware implementations, the challenge remains the realisation of an efficient processor that can handle the complexity in physiological signals and still meets the requirement for implantation in chronic applications. This can only be met with specialised hardware implementations.

Chapter 6

A Framework for Adapting Deep Brain Stimulation Based on Parkinsonian State Estimates

6.1 Introduction

Continous DBS for PD uses high frequency stimulation to ameliorate patient condition. However, this induces side effects in patients and shortens pacemaker battery life [24]. Both can be resolved using aDBS. aDBS driven by feedback signals provides an approach that optimizes clinical benefits while minimizing side effects and battery depletion [24,38]. A commonly adopted feedback signal for aDBS is LFP. LFP are used due to their correlation to patient clinical states and the ease with which they can be acquired [78, 163]. In conventional DBS, programming of stimulation parameters are done by trained clinicians [265]. Thus, aDBS techniques that imitate human reasoning into decision making could be adopted – an example of which is fuzzy control. Fuzzy control is found in numerous applications for closed loop therapy [266–268]. It has the potential to achieve a level of expertise close to (and possibly better than) human expertise in therapy modulation [269]. However, the capabilities of fuzzy control are dependent on the level of sophistication of its rules and input signal. In this chapter, state estimates are used as input to a fuzzy controller to achieve a critic-actor control policy as shown in Figure 6.1. It leverages on a machine learning model as the critic and a fuzzy controller as the actor. This individualises therapy using patient-specific state estimates which are obtained through the machine learning models. Machine learning models were selected because of their ability to create adaptable models for complex signals using statistical attributes from the signals [270]. The choice of fuzzy control was driven by their ability to provide computationally efficient and robust decision making. Consequently, as more knowledge on PD and DBS is gained, the fuzzy rules could be updated, which provides an adaptable control scheme. The scheme has the potential to be developed into a fully implantable closed-loop DBS system. The rest of the chapter is organized as follows. Section 6.2 describes the motivation for using machine learning in disease tracking. Section 6.3 details the models and metrics used in evaluating the performance of the scheme. The fuzzy controller design is explained

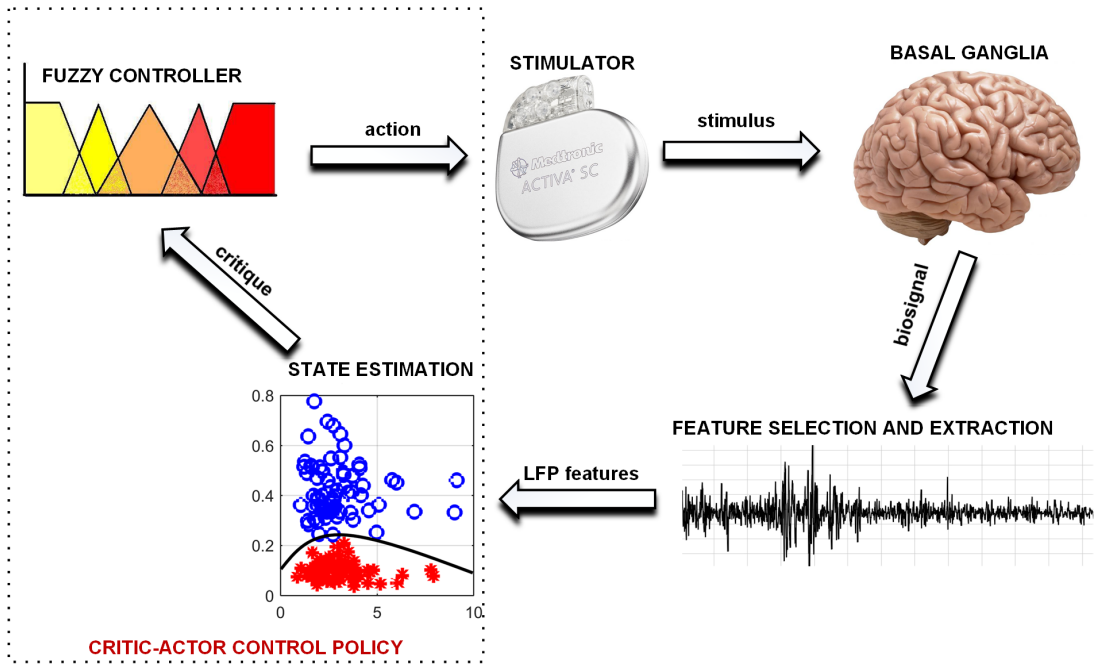


Figure 6.1: A typical scheme for adapting DBS using PD state estimates.

in Section 6.4. Performance evaluation is detailed Section 6.5. Section 6.6 outlines the prospects and limitations of the critic-actor control policy. And, Section 6.7 summarises the major points addressed in the chapter.

6.2 Machine Learning for Disease Tracking

Disease tracking is important because dynamic changes in PD pathophysiology could help inform treatment strategies. This can be achieved using machine learning models as they provide insights on disease progression. In BMI applications, machine learning provides the ability to notify caregivers of life-threatening events related to chronic disease diagnosis and management [261]. Using closed-loop control strategies, this useful information can be used to generate actionable outputs – mostly from stimulation devices – to mitigate patient conditions [271]. Machine learning models for disease tracking are intended to achieve one of two outcomes: prediction or state estimation. Prediction requires the isolation of dynamic features preceding critical events. This can be computationally intensive due to the complex and progressive nature of neural signals and the hardware requirements for tracking their real-time fluctuations. On the other hand, state estimation involves identifying critical events that could be suggestive of disease processes. For optimal delivery of bio-electronic therapy, prediction is the most desirable outcome. Nevertheless, early and accurate state estimation can be used to adjust therapy to suit patients' needs. State estimation tracks fluctuations in PD symptom severity so that stimulation can be modulated correspondingly. As such PD symptom severity is represented by the probability that a patient signal is a PD condition. A good understanding of the system dynamics is essential for achieving adequate control. The state estimator measures the PD dynamics to enable the controller

to update stimulation appropriately. The complex nature of the underlying processes in physiological signals has led to the development of various machine learning frameworks for modelling data. These frameworks give an insight into the evolution of the pathophysiological processes in patients and they help to model transitions between various states in a disease. Machine learning algorithms are classified into supervised (using labelled data), semi-supervised (using partly labelled data) and unsupervised (using unlabelled data) learning algorithms. Table 6.1 highlights the working principle of each of the classes of machine learning algorithms.

Table 6.1: Working principle of the major classes of machine learning algorithms.

Algorithms	Operating Principle	Remarks
Supervised	The provided labels enable the formation of a robust and efficient model.	Supervised learning algorithms are adopted for use in PD tracking. They are generally the preferred approach in medical applications because of the availability of medical records with clinical annotations. These annotations provide clinicians (or trained operators) proper monitoring of disease processes, which could help inform treatment strategies.
Semi-supervised	Uses a mixed approach of supervised and unsupervised learning where some of the training data are labelled and others are not.	
Unsupervised	Builds representations from unlabeled data by finding hidden structure (e.g. clustering and dimensionality reduction).	

6.2.1 Supervised Learning Algorithms

These algorithms are not only concerned with detecting patient states but can also be used in understanding the evolution of the pathophysiological processes in patients; thus modelling transitions between various states in a disease. Supervised learning algorithms are divided into discriminative and generative machine learning models.

6.2.1.1 Discriminative Algorithms

Discriminative models focus on detecting disease or non-disease states, in this case, PD and non-PD states. In BMI applications, various discriminative frameworks have been proposed, notable among are the logistic regression, support vector machine (SVM) and k-nearest neighbour (k-NN) techniques – all of which were discussed in Chapter 4. Logistic regression and SVM are parametric models while the k-NN is a non-parametric model. Generally, non-parametric models like k-NN do not require the normal iterative learning phase necessary to fit the training data to a classification model. Nevertheless, their major setback is their storage requirements.

6.2.1.2 Generative Algorithms

Generative algorithms are particularly useful in applications were the sequence of transition between states is essential in determining future states, like in sleep-stage moni-

toring applications [272]. In aDBS, they can be principally useful in applications were stimulation parameters are defined by the evolution of the sensed neural potentials. The most common generative frameworks are the Gaussian mixture model (GMM) and the hidden Markov model (HMM). HMM uses hidden (non-observable) states that can be inferred from a series of observable states to model the underlying processes. Figure 6.2 shows a graphic representation of how the hidden states are modelled in a typical generative model. In the example generative model in Figure 6.2, each state is defined using three parameters: state transition probability (T), output observation probability (O) and initial state probability (I). Figure 6.2 has two hidden states and four output observations. Each node represents a state or output observation, with the arrows representing state transition probability (t_{ij}) between the current state S_i and the next state S_j. The output probability O_{ij} depicted by the dashed arrows, represents the probability that the observation in O_j is from state S_i. The initial state probability (I_i), represents the probability that the sequence of observations start from state S_i. In Figure 6.2, state estimates are obtained using information on the output observation and the previously recorded state. This can be used to obtain the probable sequence of state of a process. Compared to discriminative algorithms, generative algorithms present a more detailed analysis of the underlying processes involved. This study will investigate the efficiency of a state estimator from each of the classes to validate the proposed framework for adapting DBS.

6.2.2 Representative State Estimators

Generative algorithms model the data based on the joint probability distributions between its classes (PD and non-PD) while discriminative algorithm models data based on their conditional probability distribution. One example in each of the two models was used to test the soundness of the proposed framework for aDBS. A linear kernel SVM was selected as the representative algorithm for the discriminative models while the Gaussian mixture model (GMM) was selected for the generative algorithms. SVM and GMM were selected because of their computational efficiency compared to other similar algorithms. Figure 6.3 shows the gradation on a features space using the conditional probability from the SVM as state estimate and the joint probability from the GMM as state estimate.

The SVM discriminating function Equation 4.30 used in classifying test cases into either PD or non-PD classes. The score in Equation 4.30 could range from negative infinity to positive infinity depending on how large the features are, with the decision boundary (demarcating point between PD and non-PD examples) represented by a score of 0. The score provides a measure of how far a test case is from the decision boundary. Due to the unbounded range of the scores, the scores are normalised to a range between 0 and 1 depending on their magnitude, with the decision boundary represented by the region having a probability of 0.5. For the feature space in Figure 6.3 (a), the obtained probabilities using the SVM discriminating function is illustrated in Figure 6.3 (c). The GMM function in Equation 6.1 provides the level of similarity between a test case and the training examples (with PD training examples as the reference point). If a test

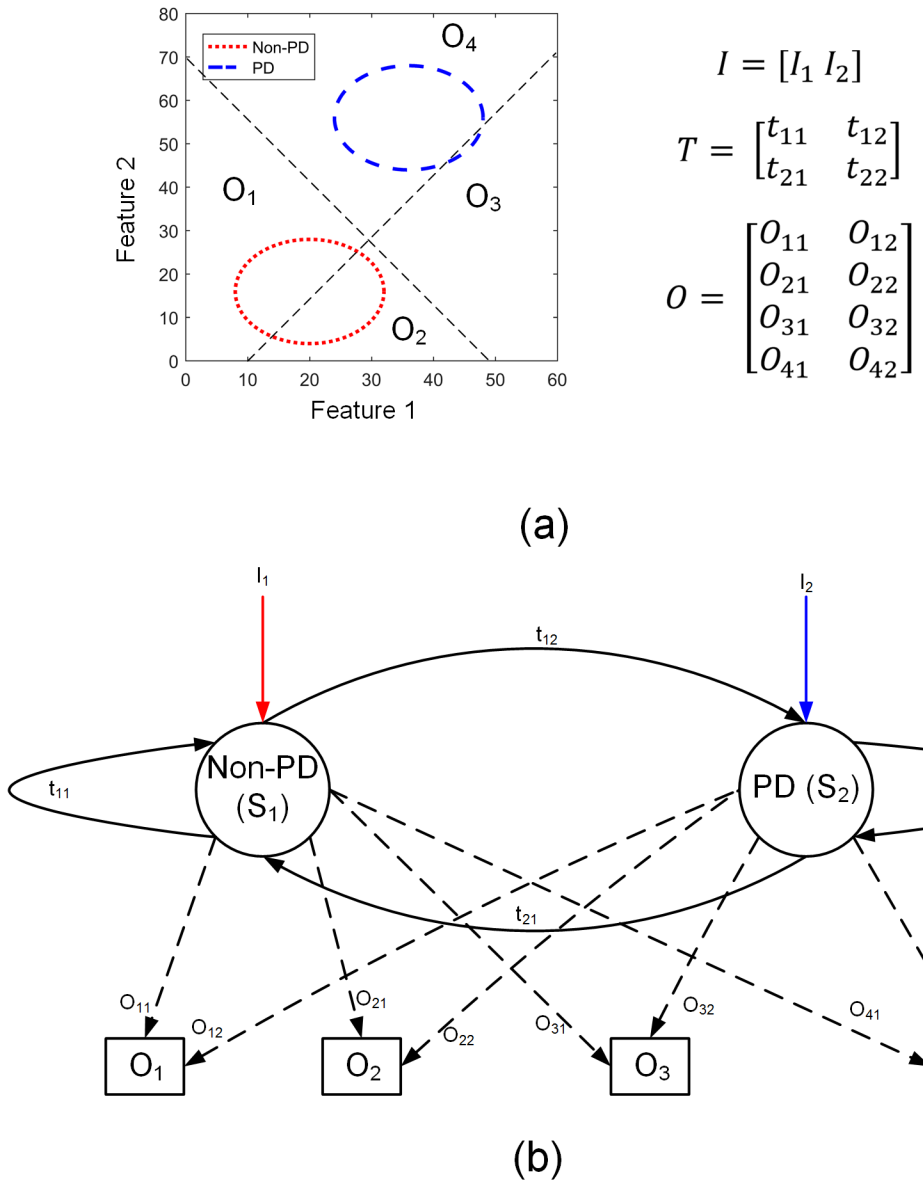


Figure 6.2: Graphical representation of a typical generative model. (a) Depicting the various output observations on the feature space and the various parameters of the model with: I as the vector of the initial probability of the states, T is a matrix of the transition probability between states and O is a matrix of the output probabilities. (b) A generative model having two hidden states (PD and non-PD) with four output observations (O_1, O_2, O_3, and O_4).

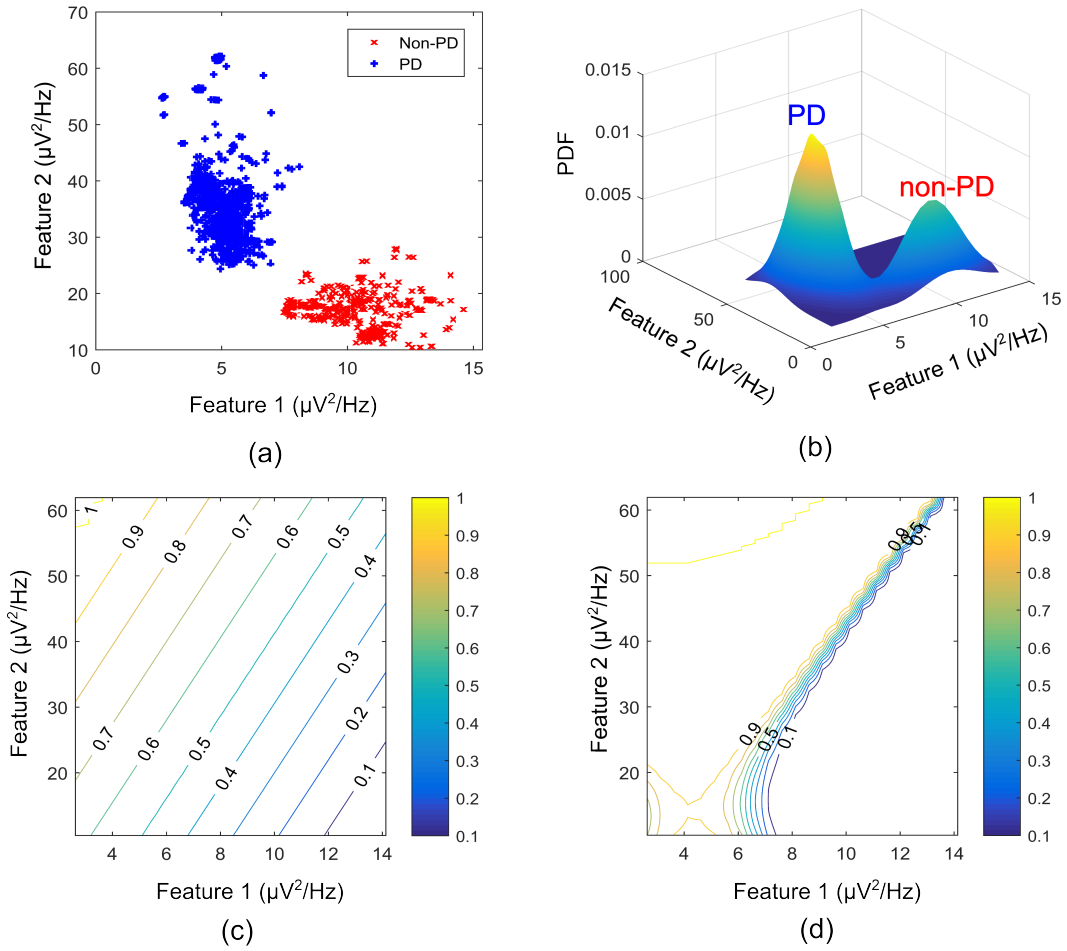


Figure 6.3: Contour plot for state estimates over a feature space for the machine learning models. (a) Feature space for dataset C. (b) Probability density function (PDF) for PD and non-PD training examples. (c) Contour plot for state estimates using SVM, with a range form 0 - 1 representing levels of severity from non-PD to PD. (c) Contour plot for state estimates using GMM, with a range form 0 - 1 representing levels of severity from non-PD to PD. The two features are, Feature 1 (21-26 Hz band) and Feature 2 (18-23 Hz band).

case presents a very good similarity to PD training examples it will have a probability approaching 1, and if it has a good similarity to non-PD training examples it will have a probability approaching 0 as can be seen in Figure 6.3 (d). Like the case for LR (in Section 4.4.3.2 of Chapter 4) probabilities greater than or equal to 0.5 are classified as PD and those less than 0.5 are classified as non-PD cases. This is why PD regions are points on the feature space where the probability is greater than 0.5, while non-PD regions are those in which the probability is less than 0.5. Thus, from non-PD to severe PD is a transition in probability from 0 to 1.

The SVM uses the widest margin between differing states to discriminate. The discriminating function for the SVM is given in Equation 4.30 as adapted from [238]. Figure 6.3 (c) depicts the state estimates obtained using SVM on the feature space in Figure 6.3 (a), whose probability density function (PDF) is shown in Figure 6.3 (b).

The GMM estimates joint probability using a weighted sum of a number of PDFs. These PDFs are used to form the Gaussian models. The weighted Gaussian functions $f_{GMM}(x)$ modeling the underlying processes are:

$$f_{GMM}(x) = \sum_{i=1}^{N} w_i \, exp\left((\overrightarrow{x} - \mu_i)^T \Lambda_i (\overrightarrow{x} - \mu_i)\right) \tag{6.1}$$

where w_i is the weight assigned to a particular Gaussian model i, $\vec{x}$ is the input feature vector for test case, μ_i is the mean vector for Gaussian model i and Λ_i is the covariance vector. The major assumption employed in GMM is that the population of feature vectors can be represented by N Gaussian models. For this application, two Gaussian models are fitted in the training data, in order to represent each of the patient states (PD and non-PD). Figure 6.3 (c) shows the state estimates obtained using SVM on the feature space in Figure 6.3 (a), whose PDF is shown in Figure 6.3 (b).

6.3 Models and Metrics

The proposed model in Figure 6.1 consists of a physiological basal ganglia network, a feature extraction stage, a state estimation stage for diagnosing PD severity and an adaptive stimulator for delivering therapy. The basal ganglia network uses LFP recordings to mimic the underlying mechanism of PD. LFP signals from the basal ganglia model are applied to a feature processing stage, and the output from this stage is applied to a regression model (state estimation stage) which estimates the patient state. Stimulation parameters are adjusted based on patient state estimates. The model was implemented using custom SIMULINK blocks (level-2 MATLAB S-function). The S-function implementation operated at 128 Hz, which is the sampling frequency of the original LFP signals. Thus, mimicking real-time operation. This was used to validate the adaptive brain stimulation scheme that restored LFP characteristics with PD condition to non-PD conditions by applying stimulation based on parkinsonian state (symptom severity).

6.3.1 Basal-Ganglia Network Model

To validate these methods, a basal ganglia model using LFP recordings obtained from measurements made on patients exhibiting a combination of bradykinesia and/or rigidity during the onset of PD, with less noticeable tremor was employed. The network which is shown in Figure 6.4, it consists of: patient LFP signals, modulating network and the modulated LFP signal.

Patient LFP Signals

These are LFP signals consisting of PD and non-PD periods synthesized from real LFP recordings. The LFP synthesis, involved fitting ARMA models to the real-life recordings to produce semi-synthetic LFP signals. Fitting an ARMA model provides the flexibility to manipulate the signal characteristics such that all underlying conditions can be represented. Also, it offers the opportunity to generate LFP signals consisting of PD

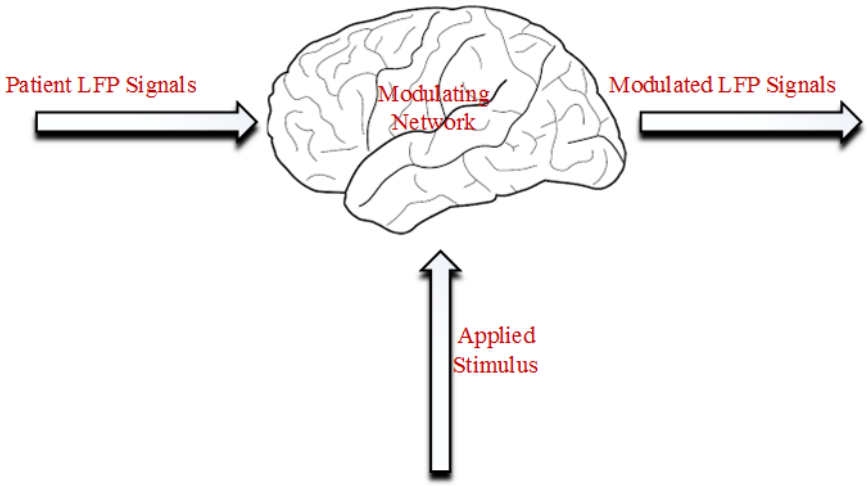


Figure 6.4: Basal-ganglia network model.

and non-PD episodes of variable duration. The complete LFP data synthesis process and a detailed description of the LFP recordings are provided in Chapter 4 and in [90]. For the aDBS framework, on detecting PD-like LFP, the applied stimulus regulates the input LFP signal such that the modulated LFP characteristics are restored to those resembling non-PD LFP. For input LFP with non-PD characteristics, the modulated LFP signal is the same as the patient LFP signal.

Modulating Network

The therapeutic mechanisms of DBS on neuronal activities are still not clear. Various studies suggest that it reduces neuronal activities [273], while others claim that it increases neuronal activities [33]. Later studies provide other alternative explanations [32]. From the studies, what is clear is that DBS has a modulating effect on neuronal activities at the stimulation site. Chapter 1 discussed a number of theories posited for DBS mechanism. In this model, a black box approach was used to model the effect of stimulation on patient LFP signals. In response to an applied stimulus, the modulating network adjusts LFP signals such that they can be restored to their PD-free state. The correlation between the changes in PD and the changes in the coefficient of variation (CV) of neuronal signals supports the hypothesis that modulating LFP signals during DBS is one of the mechanisms that can lead to PD suppression [274,275]. This prompted the two-degrees-of-freedom (2-DOF) modulation applied by the modulating network as shown in Figure 6.5. For 2-DOF modulation, the modulating network varies the patient LFP signals in the two bands with the most pronounced variation between non-PD and PD bands as shown in Figure 6.5 (b). In both bands, the headroom of variation for the magnitude of the filter response is between 1 and the ratio of CV for non-PD to PD ($\frac{CV_{NPD}}{CV_{PD}}$), as is shown in Figure 6.5 (c). This makes the modulating network unique for each patient since the frequency response of the modulating network is dependent on the relationship between PD and PD-free periods of individual patients.

Modulated LFP Signals

These are extracellular LFP signals resulting from the modification of patient LFP signals by the modulating network. The modulated LFP signals are the signals monitored by aDBS to adjust the stimulation.

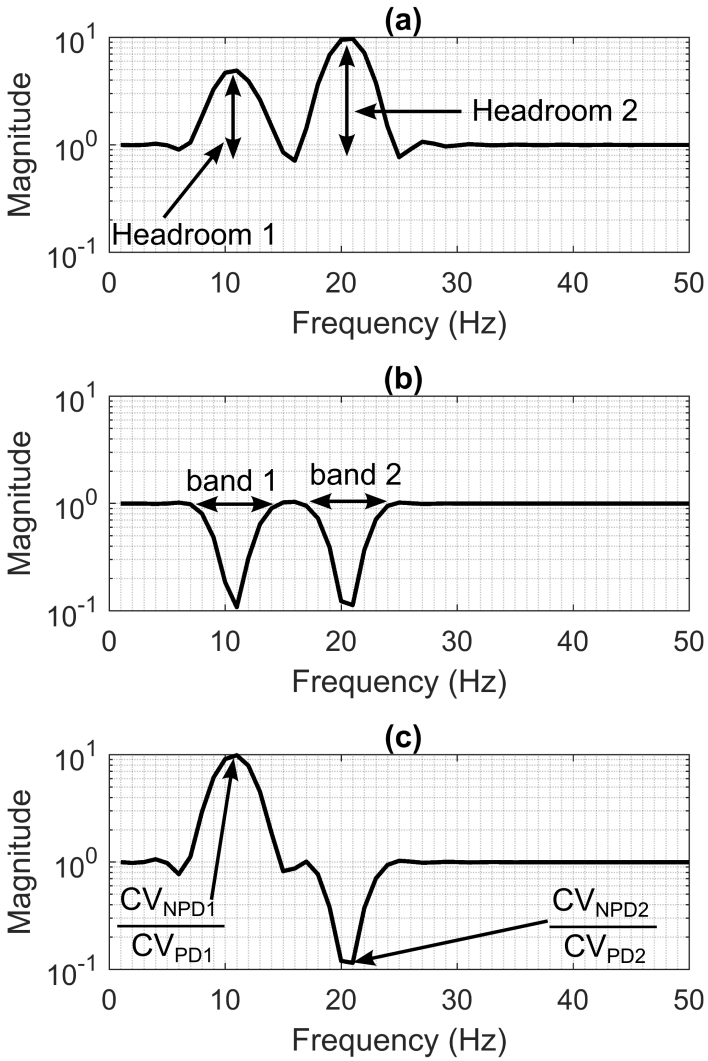


Figure 6.5: 2-DOF modulating network used to simulate the effect of DBS on neuronal signals. (a) Frequency response for configuration with non-PD having higher amplitude in both bands. (b) Frequency response for configuration with PD having higher amplitude in both bands. (c) Frequency response for configuration with non-PD having higher amplitude in band 1, and PD having higher amplitudes in band 2. CV_{NPD} is the coefficient of variation for non-PD LFP signal and CV_{PD} is the coefficient of variation for PD LFP signal.

6.3.2 Feature Extraction and Selection

6.3.2.1 Time-Frequency Analysis

It uses the FFT to obtain time-frequency data. This is achieved by dividing the signal into windows and applying FFT to each window. This was discussed in Chapter 4 and presented in Equation 4.14. For this application, the time-stamped measurements are split into 2 seconds (consisting of 256 samples) overlapping epochs, with 50% overlap between epochs. Also, the power bands (features) are divided into 5 Hz bands, with 3 Hz overlap between bands; 0 to 5 Hz, 3 to 8 Hz, . . . , 45 to 50 Hz. This provides a total of 16 features. The window is chosen such that a balance between time and frequency resolution is obtained.

6.3.2.2 Feature Selection

Feature selection involves reducing the number of features that will be used for state estimation. This is achieved using the MRM. The process of MRM starts by identifying the channel having the two bands with the most pronounced variation in activity. The goal is to obtain the bands that make state estimation easier. The MRM is a computationally simple method. It is discussed and investigated in more detail in Chapter 4.

6.3.3 Stimulation Adaptation

It is used to respond to fluctuations in the dynamics of patient LFP data. The estimated patient state is applied to the fuzzy controller and the fuzzy controller determines the appropriate stimulation parameters. The fuzzy controller is designed in Section 6.4.
The limited understanding of DBS mechanisms makes the selection of stimulation parameters (i.e. amplitude, frequency and pulse duration) difficult [276]. Experimental studies have been undertaken regarding the most beneficial stimulation parameter for the reduction in motor symptoms. Some studies suggest that there are more noticeable improvements when stimulation frequency is adjusted [277–279]. However, other studies maintain that stimulation amplitude is more critical [280–282]. More research has focused on stimulation frequency alone [109, 274, 283]. Nevertheless, the major considerations in selecting stimulation parameters are patient responses to stimulation patterns and power consumption to conserve battery life [276]. The consensus is that the most beneficial stimulation frequency occurs at 130 Hz [274,277,284]. Based on the therapeutic benefits of varying the stimulation frequency, the fuzzy rules are designed to adjust the stimulation frequency. Adjusting the stimulation frequency modifies the modulating effect in a quasi-linear fashion as is depicted in Figure 6.6. The headroom for the frequency response of the modulating network in Figure 6.5 corresponds to a stimulation frequency ranging from $0 - 90$ Hz (i.e. a magnitude response of between 1 to $\frac{CV_{NPD}}{CV_{PD}}$). This is shown in Figure 6.6, where increasing the stimulation frequency by 45 Hz moves the test case marked X to the centre of the non-PD cluster, while a decrease of 45 Hz moves it to the centre of the PD cluster. In theory, a 90 Hz change in stimulation frequency can move a test case from the centre of one cluster to the other (PD to non-PD cluster or vice versa). To accommodate for outliers, the stimulation

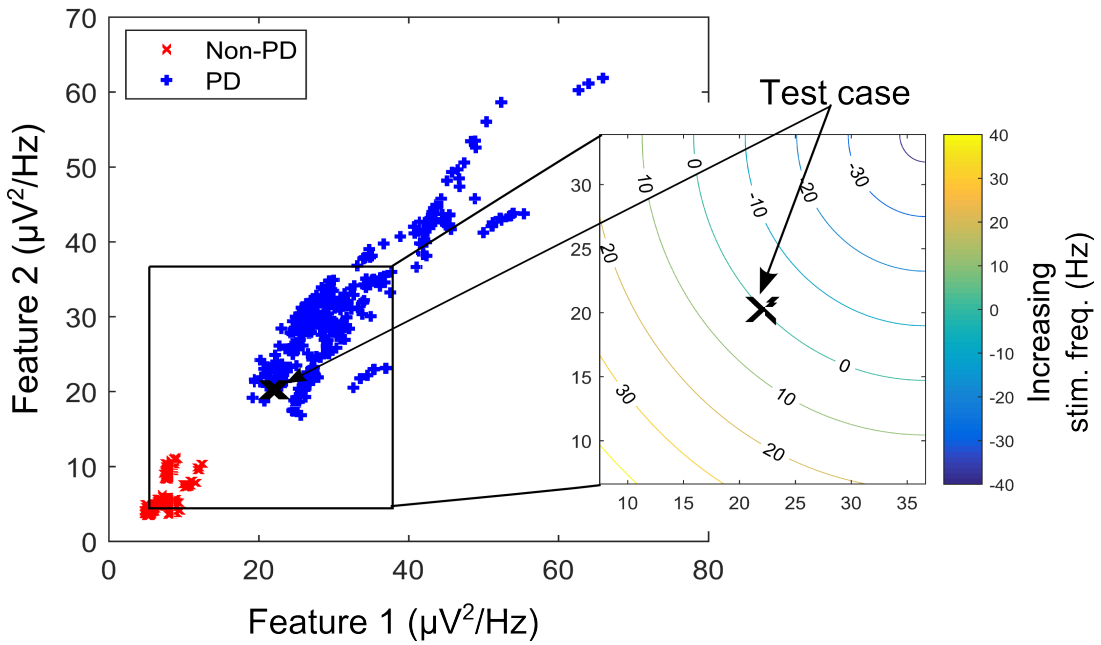


Figure 6.6: A contour plot depicting the effect of increasing stimulation frequency on the transition path of a test case (in the XY-location marked "X") over the feature space. Feature space is that of dataset B.

frequency is capped at 180 Hz, which is normally the limit for conventional DBS.

6.3.4 Evaluation Metrics

In assessing the performance of the different state estimator based approaches, three measures that are indicative of accuracy, latency and computational complexity have been used.

6.3.4.1 Accuarcy

The state estimators are evaluated using MCC and WCE, both of which were discussed in Chapter 3. MCC and WCE are balanced measures used in assessing state estimator quality that can be used even for cases with skewed classes. MCC measures the correlation coefficient between the observed and predicted binary classifications. It has a range between -1 (total disagreement) and $+1$ (total agreement); with 0 representing a random prediction. WCE is a weighted representation of type I and type II error from the confusion matrix and it can be represented mathematically as,

$$Error_{weighted} = \frac{1}{2}\left(\frac{FP}{FP+TN} + \frac{FN}{TP+FN}\right) \tag{6.2}$$

where TP are the true positives, FP the false positives, FN the false negatives and TN the true negatives. The major shortcoming of MCC is that it can only be used when one of the denominators $TP+FN$, $TP+FP$, $TN+FP$ and $TN+FN$ is not a zero. The first part of Equation 6.2 represents type I error (false-positive rate), while the second part represents type II error (false-negative rate). In Equation 6.2, WCE uses equal weights to compute the average of type I and type II error.

6.3.4.2 Latency

Detection latency in this work measures the total time required by the system (or controller) to settle to the modal class interval for non-PD state defined by the fuzzy controller. For SVM driven control policy, it is a state between 0.15 and 0.35, while for GMM, it is a state between 1×10^{-8} and 0.1.

6.3.4.3 Computational Complexity

A critical consideration in assessing complexity cost in BMI processors is determining the arithmetic operations required to decode information from neural activity. This is particularly important for resource efficient hardware implementation. Arithmetic manipulations are performed on recorded signals to transform them into useful, manageable or computationally efficient forms. In this work, the primary concern is the computational cost of the critic-actor control algorithm consisting of the state estimator and the fuzzy controller. Computational cost consists of two components, NOP and memory requirements. NOP is measured using the number of additions and number of multiplications. It can be represented mathematically as [91],

$$NOP = N_{add(sub)} + Res \cdot N_{mult(div)} \tag{6.3}$$

where $N_{add(sub)}$ is the number of 1-bit additions or subtractions; $N_{mult(div)}$ is the number of 1-bit multiplications and divisions required in each algorithm; and Res is the resolution of the data converter used. For memory estimates, the number of 1-bit registers required are obtained.

6.4 Fuzzy Controller Design

Based on the severity of parkinsonian state, fuzzy rules are defined to regulate stimulation based on parkinsonian state estimates. The fuzzy controller modifies the frequency of the stimulation applied to the modulating network to suppress PD-related oscillations. A fuzzy controller was chosen because it uses a reasoning which is similar to human reasoning and decision making. This makes it superior in handling non-linearity and uncertainty compared to schemes like proportional-integral-derivative controllers, lead-lag and state feedback control [147, 285]. Fuzzy controller design essentially involves the following:

- Choosing the fuzzy controller inputs and outputs.
- Choosing the pre-processing that is required for the controller inputs and the post-processing for the controller outputs.
- Designing the four components of the fuzzy controller (rule-table, inference mechanism, fuzzification and defuzzification as shown in Figure 6.7).

Figure 6.7 depicts the block diagram of a fuzzy controller for adapting DBS. To facilitate the design of the fuzzy controller, Figure 6.8 (a) and (b) shows the desired average profile

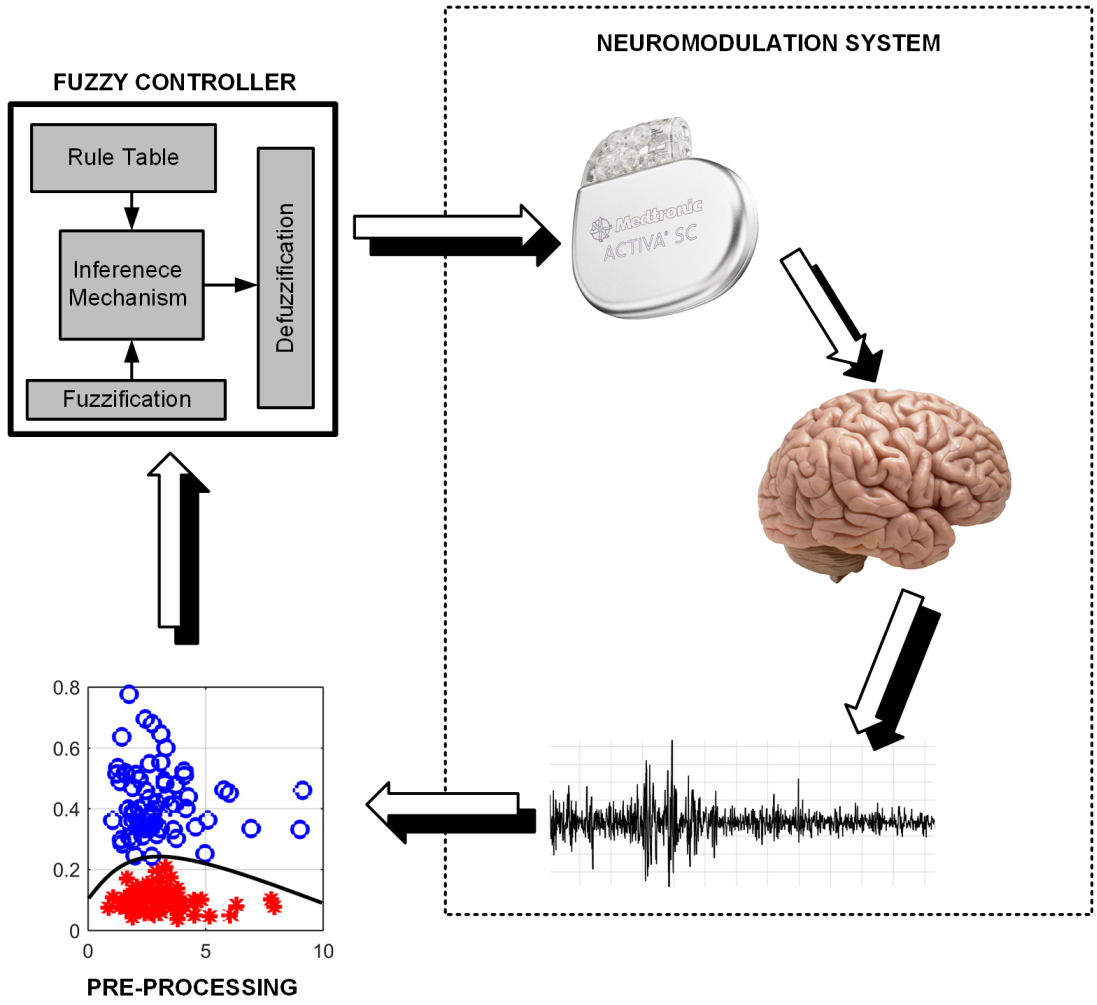
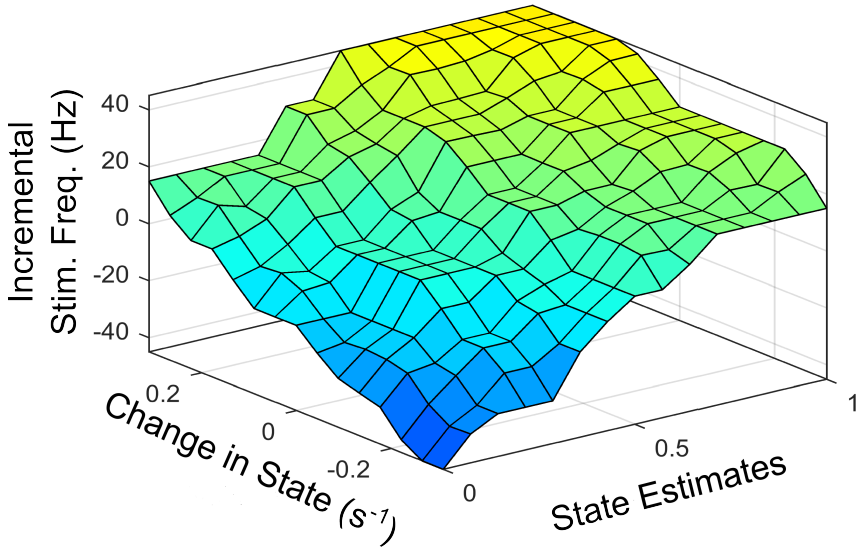


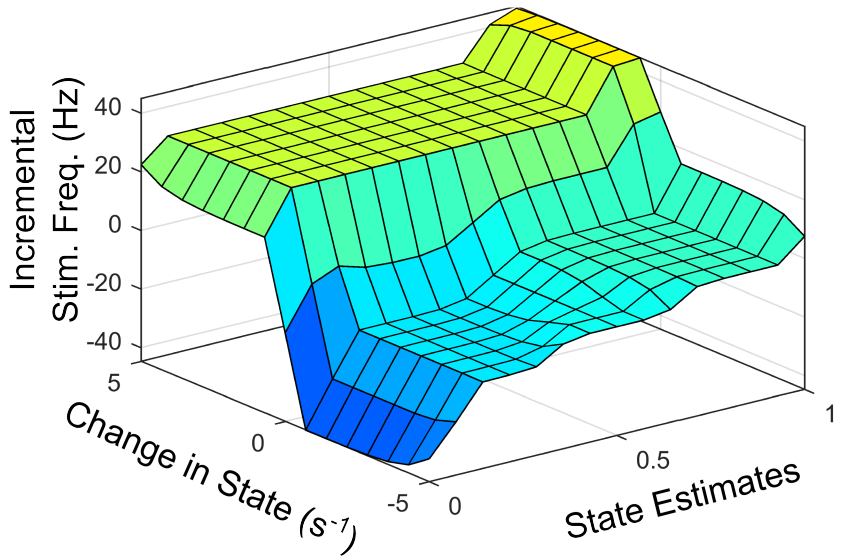
Figure 6.7: Block diagram of a fuzzy controller for adapting DBS.

for the effect of incremental stimulation frequency for all possible input combinations for the SVM and GMM driven approaches respectively.

Figure 6.8 represents the average 3-D profile that maps state estimates and change in state to stimulation frequency. To obtain the profile or each patient dataset, training examples at discrete points on the feature space representing states estimates ranging 0 to 1 in steps of 0.1 are identified. For training examples at each discrete point on the feature space, stimulation frequency is increased in steps from -45 to $+45$ Hz (in steps of 5 Hz). The corresponding rate of change in patient state is obtained for each discrete pair consisting of patient state estimate and applied stimulation frequency. This produces a mapping of three variables (patient state estimates, change in state and applied stimulation frequency). This means, for every patient state, there is an applied stimulation frequency that results in a specific rate of change in patient state. The process is repeated for all nine patient datasets and the average for the various profiles obtained as Figure 6.8 (a) and (b). For the SVM based approach, Figure 6.8 (a) represents the average 3-D profile that maps state estimates and change in state to stimulation frequency. This is represented by Figure 6.8 (b) for the GMM based approach. The average profiles in Figure 6.8 are used to guide the rule-tables used to



(a)



(b)

Figure 6.8: Surface plot for input-output relationship for: (a) SVM based controller, (b) GMM based controller.

control PD suppression. The profile for the change in state (measured in s^{-1}) targets a settling time of between 1 to 1.5 seconds from the centre of the modal class of the PD state (with a probability 0.75 for SVM, and 0.9999 for GMM) to the centre of the modal class of the non-PD state (with a probability 0.25 for SVM and 1×10^{-4} for GMM). From Figure 6.8 it is obvious that from a PD state of 1, the SVM-driven approach has a more gradual descent, while the GMM has a sharper descent at the edges, plateaus for a range of input values in which change in input only causes a slight change in stimulation frequency before it finally descends steeply. This surface plot guided the choice of membership function and rule table for the fuzzy controller, which are normally chosen heuristically. The input-output relationship was obtained using the average profile for state estimate and incremental stimulation frequency which are depicted in Figure 6.3 and Figure 6.6 respectively.

6.4.1 Fuzzification

This is the encoding step. It modifies the inputs so that they can be interpreted and compared to the rules in the rule-table. The controller inputs are converted to information usable by the inference mechanism. Obtaining a value for an input variable and finding the numeric values of the membership functions that are defined for that variable. It can also be seen as an encoding of the fuzzy controller inputs. The encoded information is used in the fuzzy inference process that begins with matching.

6.4.2 Fuzzy Rules and Membership Functions

Fuzzy rules and membership functions are obtained by studying the plant dynamics (using modelling and simulation), based on these, a set of control rules that make sense are adopted. This makes fuzzy controller design subjective and dependent on the expert designer. Also, the adaptable nature of a fuzzy controller makes it a suitable candidate, since the mechanisms of DBS are still under debate. The control scheme uses a two-input one-output fuzzy controller. The inputs are PD state estimate and the rate of change in state. The state estimates and rate of change in state quantify the dynamics of the underlying process to enable control. State estimates are obtained using SVM and GMM. The output is the incremental stimulation frequency. Based on the gradation of the state estimates in Figure 6.3 (c) and Figure 6.3 (d), and the contour plot depicting the effect of stimulation frequency in Figure 6.6, triangular membership functions were used for the inputs and output of the SVM driven approach. While for the GMM based approach, Gaussian functions were adopted. The rule table for the SVM-driven control approach is shown in Table 6.2. It is obtained using the 3-D profile in Figure 6.8 (a) representing the mapping between state estimates and change in state to stimulation frequency. By representing each triangular membership function of the state estimates by the equation,

$$\mu_A(a; x_i, y_i, z_i) = \begin{Bmatrix} 0, & a \leq x_i \\ \frac{a-x_i}{y_i-x_i}, & x_i \leq a \leq y_i \\ \frac{z_i-a}{z_i-y_i}, & y_i \leq a \leq z_i \\ 0, & a \geq z_i \end{Bmatrix} \tag{6.4}$$

where μ_A is the certainty of membership in a fuzzy set, a is the universe of discourse [0, 1] for state estimates (b is used for the change in state for SVM, and c is used for the output for SVM), x_i,y_i and z_i are scalar parameter representing points on the triangle. The left feet of the triangle is represented by x_i, the peak of the triangle is represented by y_i and the right feet is represented z_i. The input membership function for the rules in Table 6.2 are summarised in Table 6.3 and Table 6.4. While the membership functions for the output (incremental stimulation frequency) is summarised in Table 6.5.

For the GMM based-control approach, its rule table is shown in Table 6.6. It uses Gaussian membership functions, some of which are symmetrical about the mean, while others are unsymmetric about the mean. Equation 6.5 represents the membership for the state estimates that are symmetrical about the mean,

Table 6.2: Rule table for control policy using SVM for state estimation.

Incremental Stimulation Frequency		Change in state								
		B_{-4}	B_{-3}	B_{-2}	B_{-1}	B_0	B_1	B_2	B_3	B_4
State	A_0	C_{-3}	C_{-3}	C_{-2}	C_{-2}	C_{-1}	C_{-1}	C_0	C_0	C_1
	A_1	C_{-2}	C_{-2}	C_{-1}	C_{-1}	C_{-1}	C_0	C_0	C_1	C_1
	A_2	C_{-2}	C_{-1}	C_{-1}	C_0	C_0	C_0	C_1	C_1	C_1
	A_3	C_{-1}	C_{-1}	C_0	C_0	C_0	C_1	C_1	C_2	C_2
	A_4	C_0	C_0	C_0	C_0	C_1	C_1	C_2	C_2	C_2
	A_5	C_0	C_0	C_0	C_1	C_1	C_2	C_2	C_3	C_3
	A_6	C_0	C_1	C_1	C_1	C_2	C_2	C_3	C_3	C_3
	A_7	C_1	C_1	C_1	C_1	C_2	C_2	C_3	C_3	C_3
	A_8	C_1	C_1	C_1	C_2	C_2	C_2	C_3	C_3	C_3
	A_9	C_1	C_1	C_2	C_2	C_2	C_3	C_3	C_3	C_3
	A_{10}	C_1	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3

Table 6.3: Membership function for state estimates using SVM based approach.

Fuzzy Set	Membership Function (μ_A)
$A_0 = [x_0 = 0, y_0 = 0, z_0 = 0.10]$	$\mu_A(a; x_0, y_0, z_0)$
$A_1 = [x_1 = 0, y_1 = 0.10, z_1 = 0.15]$	$\mu_A(a; x_1, y_1, z_1)$
$A_2 = [x_2 = 0.15, y_2 = 0.25, z_2 = 0.35]$	$\mu_A(a; x_2, y_2, z_2)$
$A_3 = [x_3 = 0.35, y_3 = 0.40, z_3 = 0.45]$	$\mu_A(a; x_3, y_3, z_3)$
$A_4 = [x_4 = 0.40, y_4 = 0.45, z_4 = 0.50]$	$\mu_A(a; x_4, y_4, z_4)$
$A_5 = [x_5 = 0.45, y_5 = 0.50, z_5 = 0.60]$	$\mu_A(a; x_5, y_5, z_5)$
$A_6 = [x_6 = 0.50, y_6 = 0.60, z_6 = 0.70]$	$\mu_A(a; x_6, y_6, z_6)$
$A_7 = [x_7 = 0.60, y_7 = 0.70, z_7 = 0.80]$	$\mu_A(a; x_7, y_7, z_7)$
$A_8 = [x_8 = 0.70, y_8 = 0.80, z_8 = 0.90]$	$\mu_A(a; x_8, y_8, z_8)$
$A_9 = [x_9 = 0.80, y_9 = 0.90, z_9 = 1]$	$\mu_A(a; x_9, y_9, z_9)$
$A_{10} = [x_{10} = 0.90, y_{10} = 1, z_{10} = 1]$	$\mu_A(a; x_{10}, y_{10}, z_{10})$

Table 6.4: Membership function for change in state (s^{-1}) using SVM based approach.

Fuzzy Set	Membership Function (μ_B)
$B_{-4} = [x_{-4} = -0.31, y_{-4} = -0.31, z_{-4} = -0.24]$	$\mu_B(b; x_{-4}, y_{-4}, z_{-4})$
$B_{-3} = [x_{-3} = -0.31, y_{-3} = -0.24, z_{-3} = -0.17]$	$\mu_B(b; x_{-3}, y_{-3}, z_{-3})$
$B_{-2} = [x_{-2} = -0.24, y_{-2} = -0.17, z_{-2} = -0.10]$	$\mu_B(b; x_{-2}, y_{-2}, z_{-2})$
$B_{-1} = [x_{-1} = -0.17, y_{-1} = -0.10, z_{-1} = 0]$	$\mu_B(b; x_{-1}, y_{-1}, z_{-1})$
$B_0 = [x_0 = -0.10, y_0 = 0, z_0 = 0.10]$	$\mu_B(b; x_0, y_0, z_0)$
$B_1 = [x_1 = 0, y_1 = 0.10, z_1 = 0.17]$	$\mu_B(b; x_1, y_1, z_1)$
$B_2 = [x_2 = 0.10, y_2 = 0.17, z_2 = 0.24]$	$\mu_B(b; x_2, y_2, z_2)$
$B_3 = [x_3 = 0.17, y_3 = 0.24, z_3 = 0.31]$	$\mu_B(b; x_3, y_3, z_3)$
$B_4 = [x_4 = 0.24, y_4 = 0.31, z_4 = 0.31]$	$\mu_B(b; x_4, y_4, z_4)$

Table 6.5: Membership function for Incremental stimulation frequency (Hz) using SVM based approach.

Fuzzy Set	Membership Function (μ_C)
$C_{-3} = [x_{-3} = -60, y_{-3} = -45, z_{-3} = -30]$	$\mu_C(c; x_{-3}, y_{-3}, z_3)$
$C_{-2} = [x_{-2} = -45, y_{-2} = -30, z_{-2} = -15]$	$\mu_C(c; x_{-2}, y_{-2}, z_2)$
$C_{-1} = [x_{-1} = -30, y_{-1} = -15, z_{-1} = 0]$	$\mu_C(c; x_{-1}, y_1, z_1)$
$C_0 = [x_0 = -15, y_0 = 0, z_0 = 15]$	$\mu_C(c; x_0, y_0, z_0)$
$C_1 = [x_1 = 0, y_1 = 15, z_1 = 30]$	$\mu_C(c; x_1, y_1, z_1)$
$C_2 = [x_2 = 15, y_2 = 30, z_2 = 45]$	$\mu_C(c; x_2, y_2, z_2)$
$C_3 = [x_3 = 30, y_3 = 45, z_3 = 60]$	$\mu_C(c; x_3, y_3, z_3)$

$$\mu_D(d; \sigma_i, d_i) = e^{\frac{-(d-d_i)^2}{2\sigma_i^2}}, d_i - 3.7\sigma_i \leq d \leq d_i + 3.7\sigma_i \tag{6.5}$$

where μ_D is the certainty of membership in a fuzzy set, d is the universe of discourse [0, 1] for state estimates (e is used for the change in state for GMM, and f is used for the output for GMM), σ_i and d_i are scalar parameters representing the standard deviation and mean respectively of the symmetric Gaussian function. On the other hand, the unsymmetric Gaussian functions can be represented by Equation 6.6,

$$\mu_D(d; \sigma_{ia}, \sigma_{ib}, d_i) = \begin{cases} e^{\frac{-(d-d_i)^2}{2\sigma_{ia}^2}}, & d_i - 3.7\sigma_{ia} \leq d \leq d_i \\ e^{\frac{-(d-d_i)^2}{2\sigma_{ib}^2}}, & d_i \leq d \leq d_i + 3.7\sigma_{ib} \end{cases} \tag{6.6}$$

where μ_D is the certainty of membership in a fuzzy set, d is the universe of discourse [0, 1] for state estimates (e is used for the change in state for GMM), σ_{ia} is the standard deviation of the left tail of the Gaussian membership function, σ_{ib} is the standard deviation of the right tail of the Gaussian membership function, and d_i represents the mean of the Gaussian function. Its input membership functions for the rules in Table 6.6 are summarised in Table 6.7 and Table 6.8. While the membership functions for the output (incremental stimulation frequency) is summarised in Table 6.9.

The universe of discourse for the state estimates for both SVM and GMM driven approaches is [0, 1] as can be seen in Figure 6.9 (a) and Figure 6.10 (a) for the SVM and GMM respectively. The input fuzzy sets for the SVM are represented by alphanumeric

variables $A_0, A_1 \ldots, A_{10}$, and that of the GMM is $D_0, D_1, \ldots, D_7$. This means that for state estimates, the SVM driven approach has eleven fuzzy sets and the GMM driven approach has eight fuzzy sets.

For the second input which is the change in state, the fuzzy sets of the SVM driven

Table 6.6: Rule table for control policy using GMM for state estimation.

Incremental Stimulation Frequency		Change in state								
		E_{-4}	E_{-3}	E_{-2}	E_{-1}	E_0	E_1	E_2	E_3	E_4
State	D_0	F_{-3}	F_{-3}	F_{-2}	F_{-1}	F_{-1}	F_{-1}	F_0	F_0	F_1
	D_1	F_{-2}	F_{-2}	F_{-1}	F_{-1}	F_{-1}	F_0	F_0	F_1	F_1
	D_2	F_{-2}	F_{-1}	F_{-1}	F_0	F_0	F_0	F_1	F_1	F_2
	D_3	F_{-1}	F_{-1}	F_0	F_0	F_0	F_1	F_1	F_2	F_2
	D_4	F_0	F_0	F_0	F_1	F_1	F_1	F_2	F_2	F_2
	D_5	F_0	F_0	F_1	F_1	F_2	F_2	F_2	F_2	F_3
	D_6	F_1	F_1	F_1	F_2	F_2	F_2	F_3	F_3	F_3
	D_7	F_1	F_1	F_2	F_2	F_2	F_3	F_3	F_3	F_3

Table 6.7: Membership function for state estimates using GMM based approach.

State Estimates	Membership Function (μ_D)
$D_0 = [\sigma_0 = 1 \times 10^{-10}, d_0 = 0]$	$\mu_D(d) = \left\{ e^{\frac{-(a-d_0)^2}{2\sigma_0^2}}, \quad 0 \leq d \leq 3.7\sigma_0 \right\}$
$D_1 = [\sigma_{1a} = 1\times10^{-10}, \sigma_{1b} = 3.23\times10^{-9}, d_1 = 1 \times 10^{-10}]$	$\mu_D(d; \sigma_{1a}, \sigma_{1b}, d_1)$
$D_2 = [\sigma_{2a} = 3.33 \times 10^{-5}, \sigma_{2b} = 3.33 \times 10^{-2}, d_2 = 1 \times 10^{-4}]$	$\mu_D(d; \sigma_{2a}, \sigma_{2b}, d_2)$
$D_3 = [\sigma_3 = 0.1, d_3 = 0.4]$	$\mu_D(d; \sigma_3, d_3)$
$D_4 = [\sigma_4 = 0.1, d_4 = 0.6999]$	$\mu_D(d; \sigma_4, d_4)$
$D_5 = [\sigma_{5a} = 0.1, \sigma_{5b} = 3.33 \times 10^{-5}, d_5 = 0.9999]$	$\mu_D(d; \sigma_{5a}, \sigma_{5b}, d_5)$
$D_6 = [\sigma_{6a} = 3.33 \times 10^{-5}, \sigma_{6b} = 3.33 \times 10^{-10}, d_6 = 0.999999999]$	$\mu_D(d; \sigma_{6a}, \sigma_{6b}, d_6)$
$D_7 = [\sigma_7 = 3.33 \times 10^{-5}, d_7 = 1]$	$\mu_D(d) = \left\{ e^{\frac{-(a-d_7)^2}{2\sigma_7^2}}, \quad 1 - 3.7\sigma_7 \leq d \leq 1 \right\}$

Table 6.8: Membership function for change in state(s^{-1}) using GMM based approach.

Change in state (s^{-1})	Membership Function (μ_E)
$E_{-4} = [\sigma_{-4a} = 1.5, \sigma_{-4b} = 0.165, e_{-4} = -0.5]$	$\mu_E(e; \sigma_{-4a}, \sigma_{-4b}, e_{-4})$
$E_{-3} = [\sigma_{-3a} = 0.165, \sigma_{-3b} = 0.00167, e_{-3} = -0.005]$	$\mu_E(e; \sigma_{-3a}, \sigma_{-3b}, e_{-3})$
$E_{-2} = [\sigma_{-2a} = 0.00167, \sigma_{-2b} = 1.67 \times 10^{-8}, e_{-2} = -5 \times 10^{-8}]$	$\mu_E(e; \sigma_{-2a}, \sigma_{-2b}, e_{-2})$
$E_{-1} = [\sigma_{-1a} = 1.67 \times 10^{-8}, \sigma_{-1b} = 1.67 \times 10^{-14}, e_{-1} = -5 \times 10^{-14}]$	$\mu_E(e; \sigma_{-1a}, \sigma_{-1b}, e_{-1})$
$E_0 = [\sigma_0 = 1.67 \times 10^{-14}, e_0 = 0]$	$\mu_E(e; \sigma_0, e_0)$
$E_1 = [\sigma_{1a} = 1.67 \times 10^{-14}, \sigma_{1b} = 1.67 \times 10^{-8}, e_1 = 5 \times 10^{-14}]$	$\mu_E(e; \sigma_{1a}, \sigma_{1b}, e_1)$
$E_2 = [\sigma_{2a} = 1.67 \times 10^{-8}, \sigma_{2b} = 0.00167, e_2 = 5 \times 10^{-8}]$	$\mu_E(e; \sigma_{2a}, \sigma_{2b}, e_2)$
$E_3 = [\sigma_{3a} = 0.00167, \sigma_{3b} = 0.165, e_3 = 0.005]$	$\mu_E(e; \sigma_{3a}, \sigma_{3b}, e_3)$
$E_3 = [\sigma_{3a} = 0.165, \sigma_{3b} = 1.5, e_3 = 0.5]$	$\mu_E(e; \sigma_{4a}, \sigma_{4b}, e_4)$

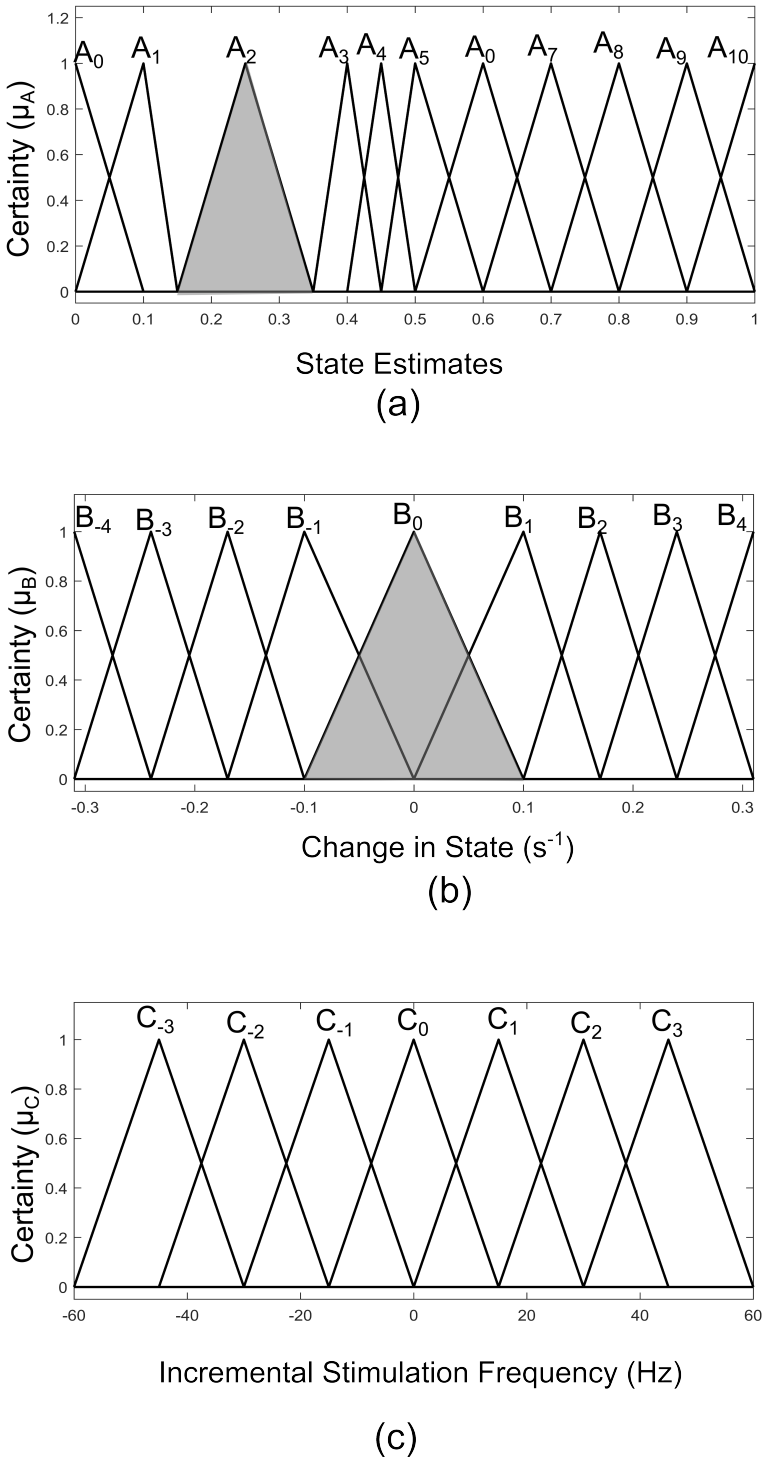


Figure 6.9: Input-output membership functions for the fuzzy controller driven by SVM state estimates. (a) Membership functions for the state estimates. (b) Membership functions for the rate of change in state. (c) Membership function for the incremental stimulation frequency.

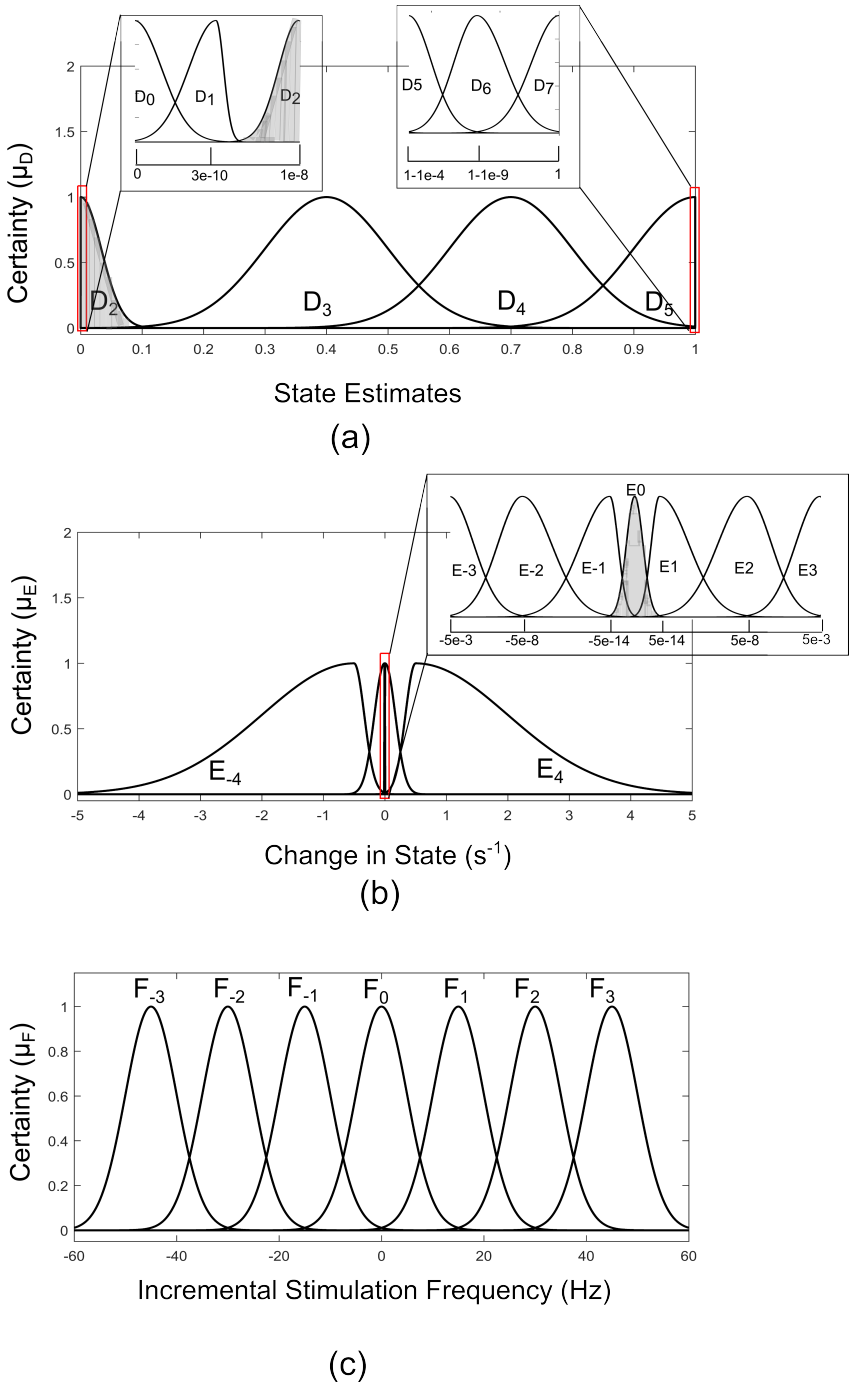


Figure 6.10: Input-output membership functions for the fuzzy controller driven by GMM state estimates. (a) Membership functions for the state estimates. (b) Membership functions for the rate of change in state. (c) Membership function for the incremental stimulation frequency.

Table 6.9: Membership function for incremental stimulation frequency (Hz) using GMM based approach.

Incremental Stimulation Frequency (Hz)	Membership Function (μ_F)
$F_{-3} = [\sigma_{-3} = 4.95, f_{-3} = -45]$	$\mu_F(f; \sigma_{-3}, f_{-3})$
$F_{-2} = [\sigma_{-2} = 4.95, f_{-2} = -30]$	$\mu_F(f; \sigma_{-2}, f_{-2})$
$F_1 = [\sigma_{-1} = 4.95, f_{-1} = -15]$	$\mu_F(f; \sigma_{-1}, f_{-1})$
$F_0 = [\sigma_0 = 4.95, f_0 = 0]$	$\mu_F(f; \sigma_0, f_0)$
$F_1 = [\sigma_1 = 4.95, f_1 = 15]$	$\mu_F(f; \sigma_1, f_{-1})$
$F_2 = [\sigma_2 = 4.95, f_2 = 30]$	$\mu_F(f; \sigma_2, f_{-2})$
$F_3 = [\sigma_3 = 4.95, f_3 = 45]$	$\mu_F(f; \sigma_3, f_3)$

approach are represented by alphanumeric variables $B_{-4}, \ldots, B_0 \ldots, B_4$, making a total of nine fuzzy sets. Their membership functions are summarised in Table 6.4. From Table 6.4, it can be seen that negative subscripts represent a change in PD probability from one towards zero (PD to non-PD) and positive subscripts represent a change from zero towards one (non-PD to PD). This is the same for the change in state of the GMM-driven approach with fuzzy sets represented by alphanumeric variables $E_{-4}, \ldots, E_0, \ldots, E_4$, and their respective membership functions summarised in Table 6.7. As summarised in Table 6.4 and Table 6.7 respectively, the universe of discourse for the SVM-driven approach is $[-0.31, 0.31]s^{-1}$ and that of the GMM-driven approach is $[-5, 5]s^{-1}$. The fuzzy sets representing the output (incremental stimulation frequency) are labelled $C_{-3} \ldots, C_3$, for the SVM approach and that of the GMM are labelled $F_{-3}, \ldots, F_3$. Like in the fuzzy sets for the change in state, the negative subscripts represent an output representing a reduction in stimulation frequency, while a positive subscript represents an output resulting in an increase in stimulation frequency. Both have a universe of discourse of [-60, 60] Hz. Based on heuristics, the SVM rule-table has an 11×9 array making a total of 99 possible rules, which are summarized in Table 6.2. For the GMM rule-table in Table 6.6, it is made up of an 8×9 array making a total of 72 possible rules.

The desired fuzzy set for the SVM driven approach is shaded in Figure 6.9 (a). The desired fuzzy set for state is between the intervals 0.15 - 0.35 [represented by fuzzy set A_2 in Figure 6.9 (a)]. This represents the modal class for non-PD cases. In terms of the change in state, the desired interval is between $-0.1s^{-1}$ to $0.1s^{-1}$ [represented by B_0 in Figure 6.9 (b)]. The modal class interval for the state estimate (A_2) was made not to overlap with other classes to avoid ambiguity in fuzzy quantification. The outermost membership functions for the inputs can be seen to saturate and values outside the range are grouped to their closest fuzzy set. However, this is not the case for the output, due to the requirement for a defined output value at any instant in time. For the GMM driven approach, the desired input values are a probability of: 1×10^{-8} to 0.1 for state estimates [represented by D_2 in Figure 6.10 (a)] and $-5 \times 10^{-14}/s$ to $5 \times 10^{-14}/s$ for change in state [represented by E_0 in Figure 6.10 (b)]. Fuzzy rules and definition of membership function are subjective and are dependent on the expert designer. That is why a wide desired range was selected in both approaches to ensure convergence. Also,

the selected range represents the modal state for stable and non-disease conditions when projected to the patient feature space, which could be demonstrative of symptom severity. The membership functions and fuzzy rules were defined carefully based on the gradation of the state estimates on the patients feature space. This was to enable a gradual and deliberate PD suppression as against abrupt and jerky response.

6.4.3 Inferenece Mechanism

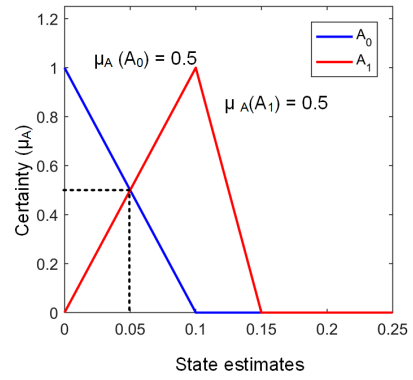
The inference mechanism generally involves two steps. Comparing the premise of all rules to the controller inputs to determine which rules apply to the current situation. It involves determining the certainty with which rules apply. The recommendations from rules that we are more certain with are adopted. Conclusions (control actions to take) are drawn using the rules that have been determined to apply at the current time. The conclusions are characterised by a fuzzy set that represents the certainty with which the input should take various values.

Premise Quantification: It involves quantifying the certainty with which our inputs belong to a fuzzy set. Then using the rules for combining premise based on Boolean logic, the certainty of the premise is obtained. There are two methods for quantifying an operation: minimum and product operation. In this work, premise is quantified using the minimum operation. Before that, the first step is to determine the applicability of each rule which is called matching. A rule is activated if its premise membership function $\mu_{premise\,(i)} > 0$. As an example, assuming the current state is 0.05 and the change in state is $-0.15\ s^{-1}$, then using the SVM rules in Table 6.2, premise quantification can be obtained as shown in Figure 6.11. Based on this, the applicable rules are,

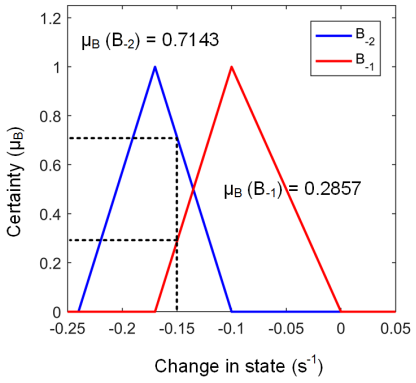
1. **if** A_0 and B_{-2} **then** $C_{-2} : \mu_{premise(1)} = min\{0.5, 0.7143\} = 0.5$
2. **if** A_0 and B_{-1} **then** $C_{-2} : \mu_{premise(2)} = min\{0.5, 0.2857\} = 0.2857$
3. **if** A_1 and B_{-2} **then** $C_{-1} : \mu_{premise(3)} = min\{0.5, 0.7143\} = 0.5$
4. **if** A_1 and B_{-1} **then** $C_{-1} : \mu_{premise(4)} = min\{0.5, 0.2857\} = 0.2857.$

Rules 2 and 4 are masked by rules 1 and 3 respectively. As a result, only $\mu_{premise(1)}$ and $\mu_{premise(3)}$ will be applicable for use.

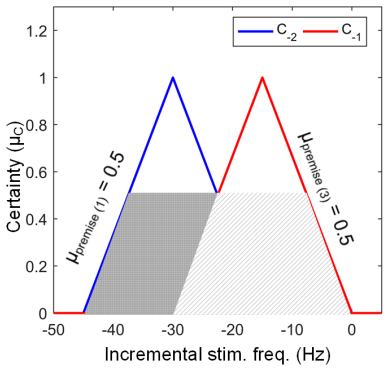
Determining Conclusions: The first step is to consider the recommendations of each rule independently. Then later the recommendations from all the rules are combined to determine the output. The minimum premise quantification is used for ANDing in above. Based on the premise quantification, the 'output fuzzy set' is determined which are C_{-1} and C_{-2}. The justification for using the minimum operator for quantification is that we can be no more certain about our consequent than the minimum of all our premise. While the input to the inference process is a set of rules that are on, its output is the set of implied fuzzy sets that represent the conclusions reached by all rules that are on. At any given time since we are using two inputs and one output we can have three possible scenarios for applicable rules: one, two or four rules can be active.



(a)



(b)



(c)

Figure 6.11: Premise quantification and defuzzification.

6.4.4 Defuzzification

This is the final operation of the fuzzy controller. It operates on implied fuzzy sets (output fuzzy sets) produced by the inference mechanism. It combines the effects of the various fuzzy sets to produce the 'most certain' controller output (plant output). Defuzzification can be considered as decoding. As the fuzzy sets produced by the inferencing process (implied fuzzy sets) is converted to numerical controller outputs. The centre of gravity (COG) method for combining recommendations represented by the implied fuzzy sets of all the rules is adopted for defuzzification. It is worth noting that:

- To avoid having an infinite area such that the COG is infinite, saturated membership functions are not allowed for the output membership functions. The COG is made not to extend beyond certain limits such that outputs above certain ranges are not applied.
- Fuzzy rules must be defined such that the sum of the denominator (in Equation 6.7) for computing COG is not equal to zero. A denominator of zero means that an undefined output will be obtained.
- Computations needed to obtain the COG of a triangle are not too significant.

If c_i denotes the centre of the membership function of implied fuzzy sets C_i, from Figure 6.11, $C_{-2} = -30$ and $C_{-1} = 15$. To obtain the additional stimulation frequency C_{stim} required to move the test case from its current position to the modal class for PD using the COG method, it can be represented as follows [286],

$$C_{stim} = \frac{\sum_i C_i A_i}{\sum_i A_i} \tag{6.7}$$

where A_i denotes the area under the membership function C_i. C_{stim} for the case in Figure 6.11, incremental stimulation frequency is obtained to be -22.5 Hz. As such stimulation will need to be increased by -22.5 Hz since the current state is 0.05 and the state is changing at a rate of $-0.15s^{-1}$. A change in state of $-0.15s^{-1}$ represents a movement away from the non-PD modal class (0.15 - 0.35 for SVM based approach). From both rule tables for the SVM and GMM driven approaches, the pattern of rule consequents show a certain symmetry. For states estimates approaching a state of 1 and having a positive rate of change in state (positive subscript, i.e. moving from non-PD to PD), there is a positive increase in stimulation frequency (positive subscript). Similarly, for state estimates approaching 0 and having a negative rate of change in state (negative subscript, i.e. moving from PD to non-PD), the incremental stimulation frequency is negative (negative subscript).

6.5 Performance Evaluation

6.5.1 PD Suppression

PD suppression is depicted in Figure 6.12 using the GMM and SVM driven approaches. From Figure 6.12 (a) which depicts SVM state estimation, it can be seen that the test case travels from the PD region and converges at the non-PD region as desired. It is also the case for the GMM approach in Figure 6.12 (c), but with a smoother trajectory. Figure 6.12 (a) and (c) show the XY-profile and Figure 6.12 (b) and (d) display the time profile.

For the time profile, it can bee seen that both cases cross the desired interval exactly

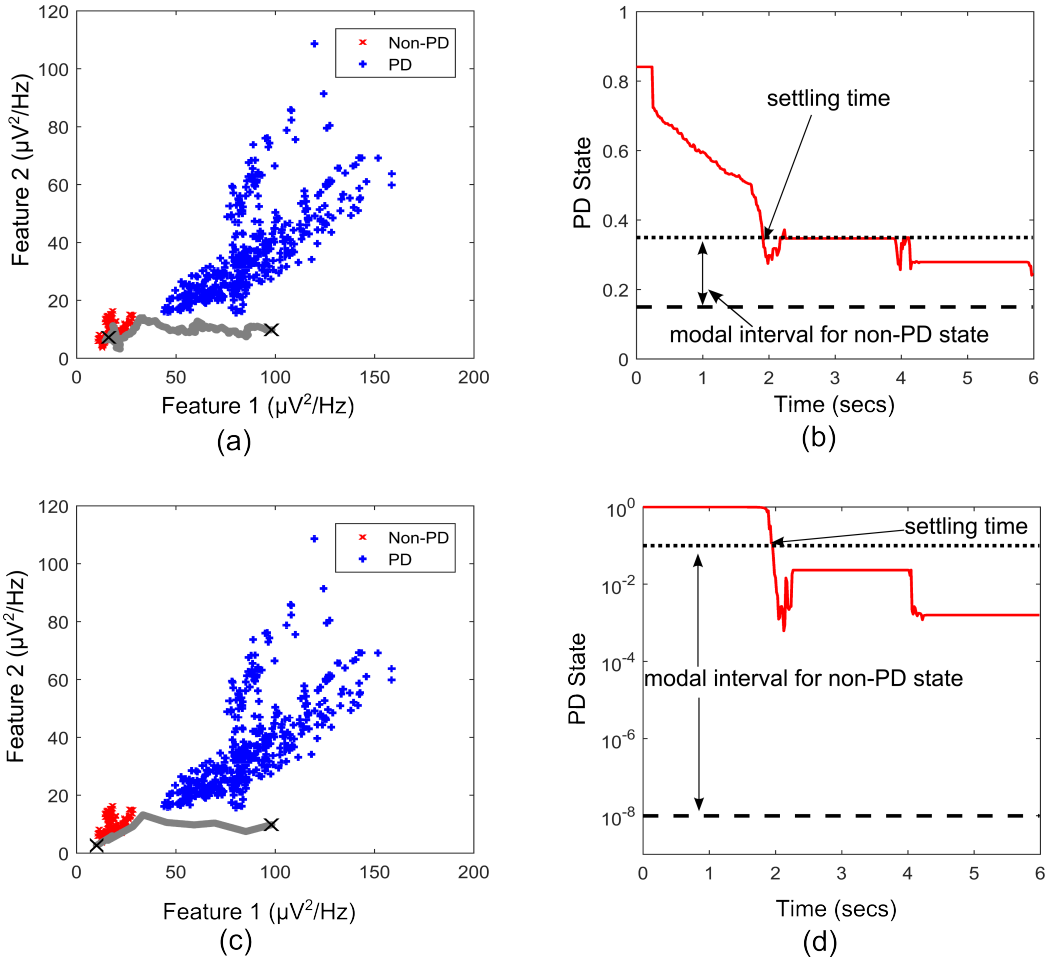


Figure 6.12: State transition of PD suppression on feature space. (a) Showing PD state transition on a feature space using SVM for state estimation, with "X" markers showing start (from PD) and settling (non-PD) positions. The XY-trajectory is indicated in grey. (b) PD state profile for PD suppression using SVM to obtain state estimates. It depicts the modal interval for the non-PD state when SVM is used for state estimation. (c) Showing PD state transition on a feature space using GMM for state estimation, with "X" markers showing start (from PD) and settling (non-PD) positions. The XY-trajectory is indicated in grey. (d) PD state profile for PD suppression using GMM to obtain state estimates. It depicts the modal interval for the non-PD state when GMM is used for state estimation. The feature space was that of dataset E.

after 2 seconds and both present the same settling profile. After settling, the SVM based approach has a mean PD state of 0.3137 and GMM-driven approach has a PD state of 1.3×10^{-2}, both of which fall within the desired range. The stimulation profile for both cases is shown in Figure 6.13. Both cases present almost the same stepwise pattern, with the SVM having a more gradual ascent to the required stimulation frequency compared to the GMM which overshoots before finally settling. The settling frequency for both cases are not far apart. The XY-profile on the feature space and the time profile (both in Figure 6.12) display a stable PD suppression profile. In addition, the stimulation profile in Figure 6.13 also displays a stable stimulation profile. Both of these are indicative of a stable PD suppression.

For the rest of the datasets, Figure 6.14 and Table 6.10 summarises their mean PD state and settling time. For the mean PD state in Figure 6.14 (a), the SVM has a lower quartile of 0.2514 and an upper quartile of 0.3162, which both fall within the desired range (0.15 - 0.35). For the GMM, it has an upper quartile of 0.0851 and a lower quartile of 2.5×10^{-4}, which are both within the desired range ($1 \times 10^{-8} - 0.1$). For the average settling times in Figure 6.14 (b), the SVM-driven approach has a lower quartile of 1.25 secs and an upper quartile of 1.875 secs. The design targeted $1 - 1.5$ seconds settling time, this is not far off the desired mark.

While for the GMM, it has a lower quartile of 0.25 and an upper quartile of 1.75 sec. The GMM based approach is a bit off the desired $1 - 1.5$ seconds settling time. Nevertheless, it has a median settling time of 1.25 seconds which falls within the desired range of values (between 1 - 1.5 second settling time). The behaviour of the GMM based-approach can be attributed to the gradation of the GMM based state estimates, which makes smooth control of its trajectory more difficult than the SVM based approach. This shows that on average, the GMM based approach has more fluctuations in settling times compared to the SVM based approach as shown in Table 6.10.

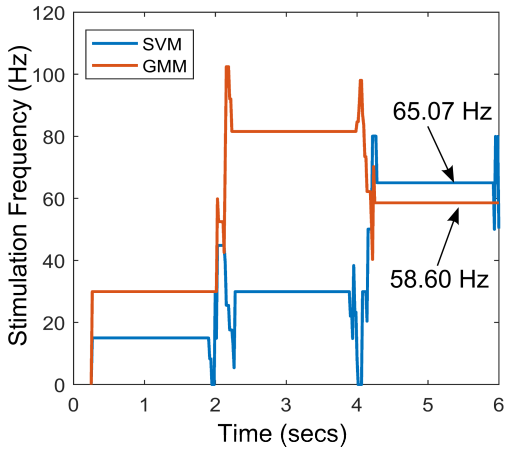
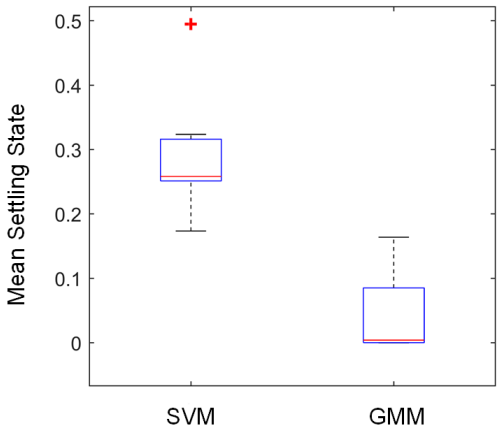
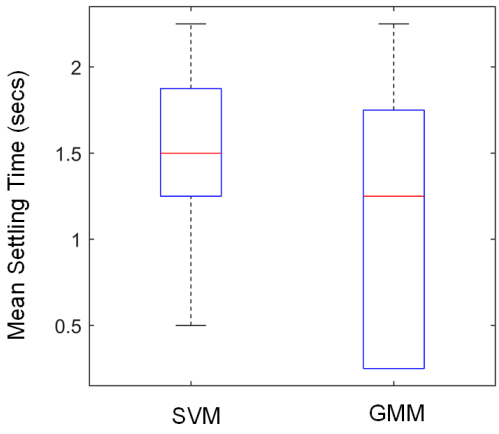


Figure 6.13: Stimulation profile for the state transition shown in Figure 6.12.



(a)



(b)

Figure 6.14: Summarizing the mean settling state and settling time for the various dataset using SVM and GMM driven approaches. (a) Mean settling state for SVM (median = 0.2584) and GMM (median = 4×10^{-3}). (b) Settling time for SVM (median = 1.5 secs) and GMM (median = 1.25 secs).

Table 6.10: Average settling time and settling state for various patient datasets.

Datasets	Average settling time (secs)		Average settling state	
	SVM	GMM	SVM	GMM
A	1.25	0.50	0.3237	0.0034
B	1.50	1.75	0.2584	0.1640
C	1.50	1.25	0.2802	3.5×10^{-4}
D	1.25	0.25	0.2547	4.5×10^{-9}
E	1.75	1.75	0.3137	0.0130
F	1.75	1.75	0.2542	0.0720
G	2.25	2.25	0.4950	0.1245
H	2.25	0.25	0.1735	4.4×10^{-20}
I	0.5	0.25	0.2431	0.0042

6.5.2 Performance of State Estimators

To assess the quality of the SVM and GMM state estimators, the MCC and WCE which are skew insensitive measures were used. The MCC measured the correlation coefficient between the original dataset and the models fitted using each of the state estimators. On the other hand, the WCE consisted of weightings of type I and type II error. This was because in aDBS, high false positive-rate will result in administering stimulation when it is not required, and this may lead to stimulation induced side effects [217]. High false-negative rate will result in the non-administering of stimulation when it may be required, which could worsen patient condition [37]. The real-time detection performance of the state estimator was investigated. Both models used 128 training examples and PD events were detected from two-second overlapping epochs (with 50% overlap). Table 6.11 summarises the average result obtained for each dataset for 100 Monte Carlo runs using 256 test cases (256 LFP epochs).

For MCC in Table 6.11, both state estimators present a positive correlation for all datasets, with the SVM having a median of 1 and the GMM with a median of 0.9433. Of the 9 cases, both SVM and GMM have 7 cases with strong positive correlation ($MCC \geq 0.5$). Only the state estimates of dataset G have a weak positive correlation in both cases. This is due to the high overlap between its PD and non-PD clusters which makes it difficult to fit the classifier to the data. From the MCC results, it can be seen that SVM fits the data better than the GMM. Similarly, the WCE results present a superior performance of the SVM over the GMM. The SVM presents a mean and median WCE of 9.03% and 0% respectively. While the GMM presents a mean and median of 11% and 1.98% respectively. This further confirms the superiority of the SVM over the GMM in fitting the data.

6.5.3 Relative Complexity

To obtain complexity estimates for both approaches, 128 training examples were assumed to be used with 8-bit quantisation (GMM inputs to fuzzy controller were assumed to have 32-bit quantization due to their resolution requirements) and 10% of the training examples were assumed to be support vectors of the SVM. The relative complexity between the SVM-driven and GMM-driven approach for each of the two stages of the

Table 6.11: State estimation performance of SVM and GMM on various patient data.

Datasets	MCC		WCE	
	SVM	GMM	SVM	GMM
A	0.3534	0.5273	0.3447	0.2204
B	1	0.8863	0	0.0771
C	1	1	0	0
D	1	0.9976	0	0.0016
E	1	1	0	0
F	0.9433	0.9433	0.0198	0.0198
G	0.4479	0.2347	0.3273	0.3757
H	1	0.9963	0	0.0012
I	0.7371	0.4343	0.1210	0.2943

critic-actor control policy are shown in Figure 6.15 . From Figure 6.15 (a), it can be seen that at the state estimation stage the SVM-driven approach requires more NOP (about 1.43 million 1-bit additions), with the GMM approach requiring only about 5% SVM NOP. At the state estimation stage, computation in the GMM is dominated by memory while for the SVM it is dominated by NOP. The GMM based approach will require about 42,352 1-bit registers to enable state estimation. This is because the GMM is a population dependent algorithm, while the SVM only uses the footprint from the population to infer properties.

In Figure 6.15 (b), the GMM requires a higher NOP for fuzzy inferencing due to its adoption of Gaussian functions as against the triangular function used by the SVM – where triangular COG is simpler to calculate. The GMM requires at least 784 1-bit additions for fuzzy inferencing as against the 208 required by the SVM based approach due to its triangular membership functions which makes COG computation easier. In terms of memory, the GMM requires fewer rules compared to the SVM. The SVM based

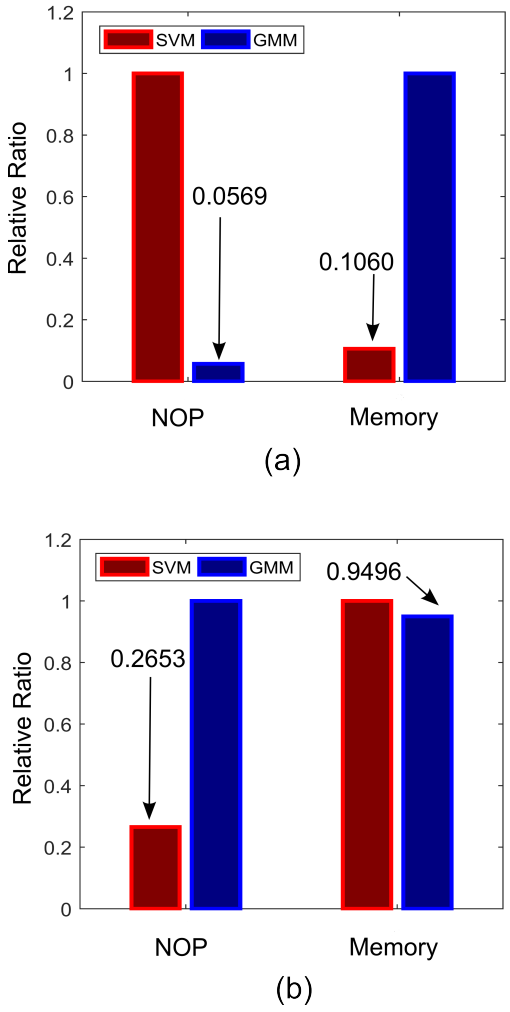


Figure 6.15: Relative complexity of the critic-actor control driven by GMM and SVM. (a) Normalized complexity for the state estimation stage. (b) Normalized complexity for the fuzzy control stage. Normalized to the maximum for all cases (maximum = 1).

approach requires at least 952 1-bit registers to store its fuzzy rules, while the GMM requires 904 1-bit registers. It is clear that in the state estimation stage the GMM has less computation and more memory, while at the fuzzy control stage the reverse is the case.

6.6 Discussion

6.6.1 Critic-Actor Control Policy

The *critic-actor* approach models the relationship between the physician and the automated neuromodulation system. The critic like the 'trained clinician' assesses the state of the system based on a cost function (in this case state estimates) and provides the information to the actor. The actor provides control signal based on evaluation from the 'informed critic'. In this configuration the state estimator is the critic, while the fuzzy controller is the actor. The main motivation for adopting the critic-actor control policy is because PD suppression can be extremely difficult to achieve due to the limited understanding of the mechanisms underlying PD. This makes it difficult to produce an accurate model that could be used for controller development. It is for this reason that more heuristic methods are proposed. The adaptive scheme exhibited the ability to restore patient LFP characteristics to PD-free conditions for different patients without a change in controller parameters. Changing conditions were monitored through the state estimates, which was the feedback signal. The feedback-loop consists of parkinsonian state (representing symptom severity) determination and stimulation facilitated by the fuzzy controller. The control signal modulates the spectral features to match PD free conditions of each individual patient. The resulting spectral features show that the adaptive scheme has the capacity to restore PD signals to their primary oscillations present under PD-free conditions. More so, using fuzzy inference mechanisms to quantify the dynamics of PD can be very intuitive for modulating therapy. Since it uses rule-based decision making that combines human heuristics into decision making; these rules could be updated into the controller as more knowledge regarding PD is acquired. Effective fuzzy control can only be achieved by adopting the right input preprocessing, in this case state estimates and their rate of change over time were chosen (as this determined the speed of PD suppression). In the future, external signals e.g. accelerometer activity can be incorporated to produce comprehensive rules that cover an increased number of possible situations. As things stand, optimal control can only be achieved by having a deeper understanding of the underlying mechanisms of DBS and PD – which is more of a clinical challenge. Ultimately, this tool could provide a paradigm on which stimulation can be adapted. The study provides a framework in which DBS can be adapted using heuristics. To validate the efficacy of the approach, state estimates were obtained using both generative and discriminative machine learning models. Both showed promising results, which are attributable to their self-calibrating nature resulting from their standard gradation on the patient feature space.

6.6.2 Model Limitations

At present, a model representing all possible dynamics is far from being realised because there is insufficient knowledge to produce models which closely represent the expected behaviour of the system. This is why PD symptom severity is represented by the probability that a patient LFP signal is a PD condition (ranging form 0 to 1). Apart from clinically sound PD state estimates, several other issues are necessary to achieve efficient PD onset control, such as optimal stimulation parameters and how they vary across patients and time. More specifically, the study focused on modulating DBS frequency; it is still under debate which of the parameters (stimulation intensity, pulse width and frequency) is the most beneficial. Nevertheless, controlling one of the parameters could shed more light on how best to control therapy. Currently, some assumptions regarding the effect of stimulation on neuronal signals are used to create a stimulation model that draws on the common denominator in all of the theories in [32,33,273]; which suggest a modulating effect on neuronal signals. This model could be improved if more detailed information on experimental LFP data consisting of stimulation parameters and PD symptom severity are obtained. Achieving significant progress in aDBS will depend on the correlation between patient state and LFP signal, as well as how stimulation modulates patient LFP. This would require a large LFP dataset representing the effect of stimulation on the progression in PD symptoms for a wide range of patients. Presently, the major challenge in adaptive DBS is the difficulty in establishing a direct relationship between patient state and stimulation parameters. This is mainly due to the complexity of post-surgery programming of stimulation parameters by trained clinicians, which can take up to six months or more [287]. Because of the limited availability of PD data incorporating the effects of stimulation, stimulation was modelled only by varying stimulation frequencies. This was chosen because stimulation frequency has proven to be more beneficial and reliable than other stimulation parameters [109,274,283].
Finally, the control policy proposed tends to work better on cases with separable classes and more granular states. For non-binary clusters (like the XOR classification problem) or binary clusters with large overlap (like the case shown in Figure 6.16), additional input information may be required to enable convergence. In Figure 6.16 the interquartile range (which is the modal class) of the PD and non-PD clusters have near complete overlap. This will make fitting a machine learning model very difficult. Convergence of the state estimates to the modal interval of the non-PD state can only be guaranteed for feature spaces with binary clusters and machine learning models that produce an MCC greater than 0.5.

6.7 Chapter Summary

The major hypothesis advanced by this chapter is that since obtaining realistic models to design controllers for the brain may be difficult, heuristic methods could be adopted to elucidate the mechanisms of DBS in patients. The following conclusions can be drawn:

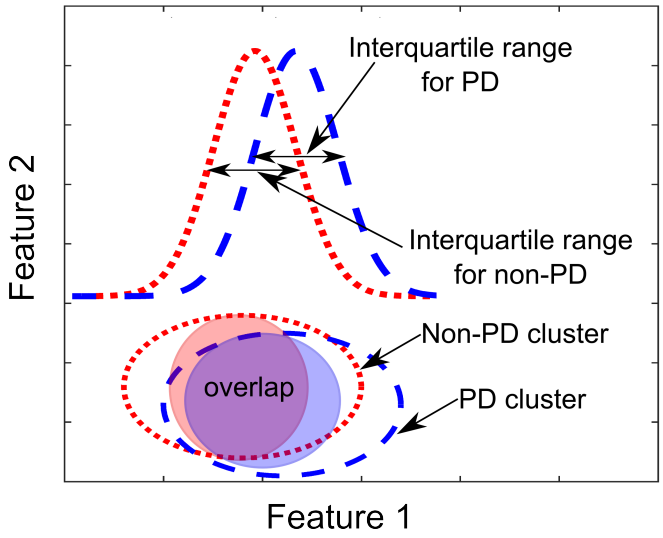


Figure 6.16: 2-D representation of feature space with a very high similarity between PD and non-PD clusters. It has a near complete overlap in the training examples that fall within the interquartile ranges of both clusters. The interquartile range of the non-PD cluster is represented by the light-red circle inscribed in the non-PD cluster, and also indicated is the PDF for the non-PD cluster. While that of PD is the light-blue circle inscribed in the PD cluster, and also indicated is the PDF for the PD cluster.

a) The dynamic progression of neural signals in PD patients necessitated the adoption of machine learning models for tracking PD. The machine learning models were used to represent the probability that an epoch of LFP signal is a PD event. These estimates were indicative of symptom severity

b) The fuzzy control approach was adopted for computational efficiency and robustness to nonlinearity. This was done with hardware implementation in mind so that the architecture can be deployed in fully implantable aDBS systems that automatically adjust stimulation parameters in real-time in response to neurophysiological signals.

c) Fuzzy rules and membership functions are obtained by studying the plant dynamics using modelling and simulation, based on these, and other requirements set by the expert designer, a set of control rules that make sense are adopted. This makes fuzzy controller design highly dependent on its designer.

d) The SVM and GMM-driven approaches used different fuzzy rules and membership functions because of their differences in grading PD and non-PD severity.

e) Both control configurations (using either SVM or GMM machine learning models) achieve PD suppression to the desired state in seven out of nine cases. This presents evidence on the possibility of mitigating intractable Parkinsonism by adaptively regulating stimulation using heuristic methods. However, it puts to question the use of universal controllers for controlling DBS as different patient datasets resulted in different settling times, settling states and state estimate profile.

f) Finally, this study presents a framework in which a control approach representing

the relationship between the trained clinician and the neuro-modulation system could be introduced into DBS adaptation. With better refinement and access to more clinical recordings, the feasibility of this approach could be explored further to ascertain if it could be deployed for clinical applications.

Chapter 7

Output Stage of a Dynamic Current Steering Stimulator for Adaptive Deep Brain Stimulation

Clinical DBS uses 4 cylindrical electrodes driven in monopolar or bipolar configurations. This yields a spherical stimulation field around the electrode modulating both targeted and untargeted areas. This chapter presents an approach in which activation field spread beyond targeted areas can be mitigated using dynamic current steering. Dynamic current steering uses segmented electrodes as in [8] together with dynamic current sources in order to accurately stimulate intended neural tissue. This chapter will focus on the output stage of the dynamic current steering system. The chapter is organised as follows: Section 7.1 gives a brief motivation for implementing the dynamic current steering system, Section 7.2 describes the high level architecture of the proposed output stage, Section 7.3 presents the simulated performance of the output stage, Section 7.4 gives a brief discussion and Section 7.5concludes the chapter.

7.1 Introduction

Cylindrical electrodes provide poor control over the stimulation field, since the resulting potential distributes symmetrically around the contact electrode and not to specific targets. This stimulates unintended neural structures and thus results in side effects [130]. As a remedy, techniques that enable accurate targeting of the electric field to the intended neural structures have been proposed [288]. These have been achieved using novel current steering techniques and novel lead designs. In terms of lead design, segmented electrodes were used in [8]. This produces radially distributed field around the electrodes and enables better targeting of neural structures compared to cylindrical DBS leads. More so, alternative methods using novel multipolar current steering stimulators were presented in [127], with [128] using both multipolar and multisite current steering. Additionally, precise neural targeting has been reported to improve the therapeutic window of DBS by reducing the threshold for beneficial effect and increasing the threshold for side effects [125]. This reduces the need for high amplitude stimulation. Current steering can also improve the long term adaptation of the system by tack-

ling lead migration in chronic deep brain stimulation. In order to achieve this, novel lead designs were proposed in [8]. In addition, the prospects of current steering led to the commercial development various current steering stimulators, notable among is the Vercise current steering system marketed by Boston Scientific. Most of the commercial developments use multiple independent current sources. This approach is uneconomical in terms of area and power consumption. In order to reduce power consumption, area and mismatch; dynamic current steering is proposed. Particular focus is placed on the output stage which uses dynamic current sources (DCS) in achieving multichannel current steering. Mismatch as well as power consumption are reduced by using a single bias voltage for both anodic and cathodic stimulation. Dynamic current sources are used in implementing 8 push-pull current sources which are interfaced to 16 electrodes through a multiplexing stage. This ensures that the output stage lends itself well to high neural selectivity and stimulation focus.

7.2 Output Stage Architecture

The architecture of the output stage is shown in Figure 7.1. It consists of a current DAC, a PMOS and NMOS DCS as in [289], binary-weighted current sources and a multiplexing stage. The output stage operates in two phases; programming (ϕ_1) and stimulation (ϕ_2). During the programming phase, the current DAC (I_{DAC}) is used in setting the bias voltage (gate-source voltage) of the DCS in Figure 7.1.
Each DCS consists of a storage capacitor and a transistor in a voltage controlled resistor (VCR) configuration as in [290]. The current DAC determines the amount of bias voltage to be stored in the gate to source capacitors of each DCS transistor, this stored voltage biases the DCS during the stimulation phase. During stimulation, the DCS

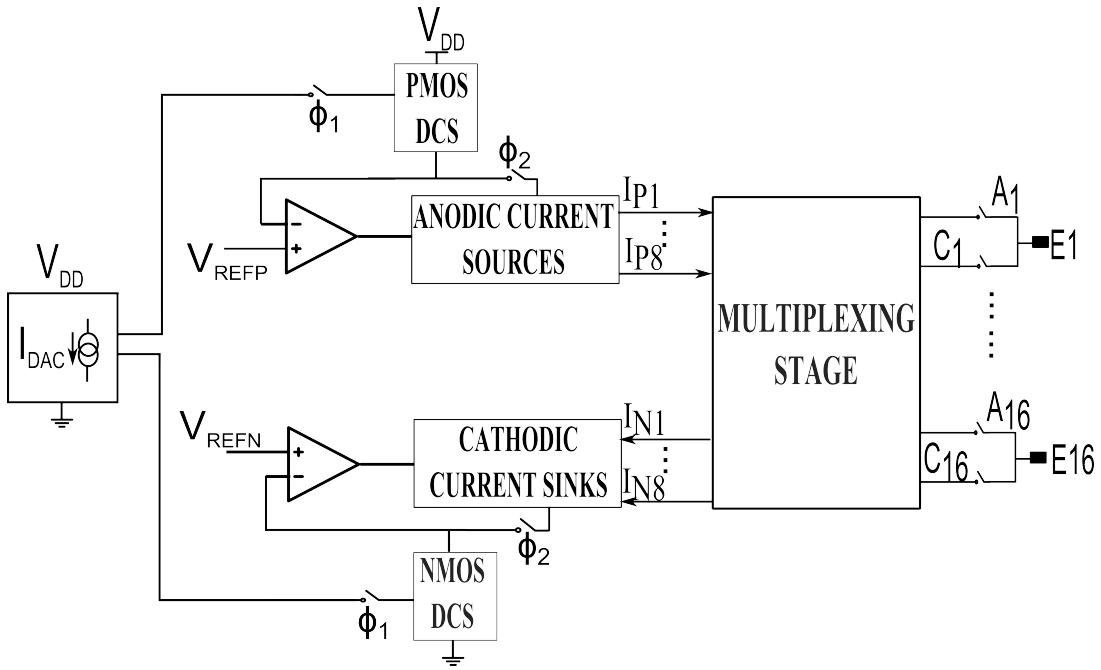


Figure 7.1: Architecture of the stimulator output stage.

routes the stimulation current through the anodic and cathodic current sources, which use binary weighted transistors in order to produce binary weighted currents, thereby creating independently weighted current sources. These binary weighted currents are steered to individual electrodes via a multiplexing stage, which provides flexibility by enabling the use of each electrode as either an anode or a cathode. The electrode configuration determines if an electrode sources or sinks current. For a proof of concept of the approach, the design parameters used are summarised in Table 7.1. Programming of the current sources adopted in this work enables the use of a single bias voltage during both anodic and cathodic stimulation phases. Moreover, it ensures that only a single branch is used for biasing and stimulation. This together reduces mismatch and power consumption. Finally, the stimulator is driven in an idle mode – with the output stage isolated from supply - during periods outside programming or stimulation.

7.2.1 Dynamic Current Sources

The DCS in Figure 7.2 is implemented using a VCR topology [289]. This was chosen in order to boost the output impedance and lower the headroom voltage. A PMOS configuration is used for the high voltage (HV) side, and an NMOS configuration is used for the low voltage (LV) side. The drain to source voltage of M_4 is set by the P_{AMP} amplifier, and that of M_1 is set by the NAMP amplifier. Both amplifiers are two stage amplifiers, with the N_{AMP} having an NMOS input stage and the P_{AMP} having a PMOS input stage. They use a bias current of 1 μA each. The amplifiers are used in a configuration such that a single amplifier is shared between programming and stimulation phases.

In order to operate the M_1 and M_4 transistors in a VCR configuration, control is required over the gate to source voltage (VGS) of each transistor. This serves as the bias voltage, and is achieved during the programming phase, by activating the programming switches in Table 7.2. The clock signals are shown in Figure 7.3. Figure 7.3 (a) shows the programming phase and Figure 7.3 (b) shows the stimulation phase. This creates a path for the bias current to the PMOS and NMOS DCS. This current flow through the DCS enables the storage of corresponding bias voltages in the gate to source capacitance of each DCS (C_{GSP} and C_{GSN}). During this process, the binary weighted push-pull current sources are isolated from the DCS using switches S_{Is1} and S_{Is2}. Afterwards, the HV and LV DCSs source and sink stimulation current through the binary weighted push-pull current sources by activating all the stimulation phase

Table 7.1: Output stage parameters.

PARAMETER	VALUE
$V_{DDH}, V_{REFP}, V_{REFN}$	3.3V, 3.25 V, 50mV
$(W/L)_P\ of HV DCS$	M_1 =185 $\mu m/1\mu m$ at 0.5 mA
$(W/L)_N\ of LV DCS$	M_4 =56 $\mu m/1\mu m$ at 0.5 mA
C_{GSP}/C_{GSN}	1 pF
Current amplitude	10 μA – 0.5mA
Pulse shape	Rectangular, biphasic

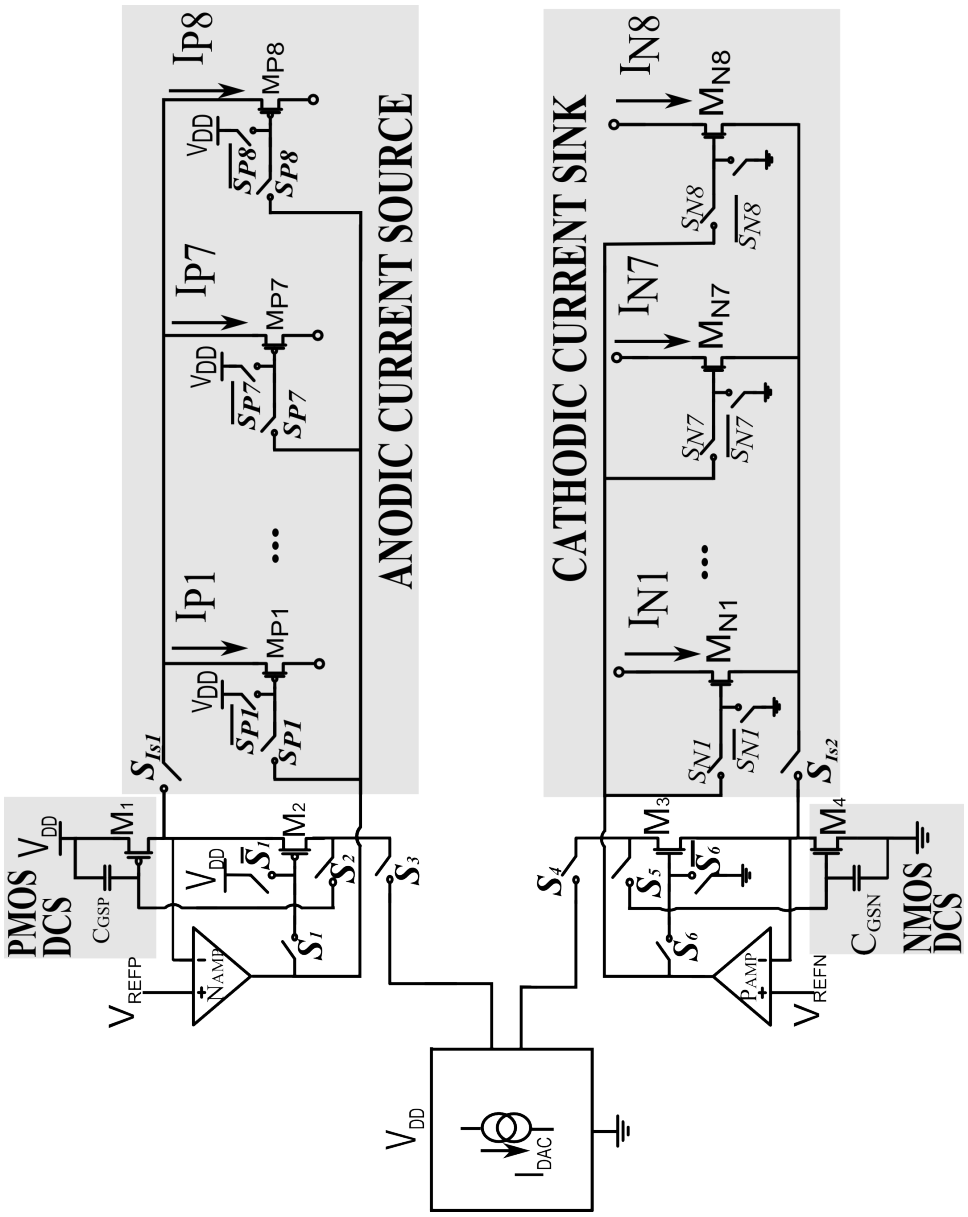
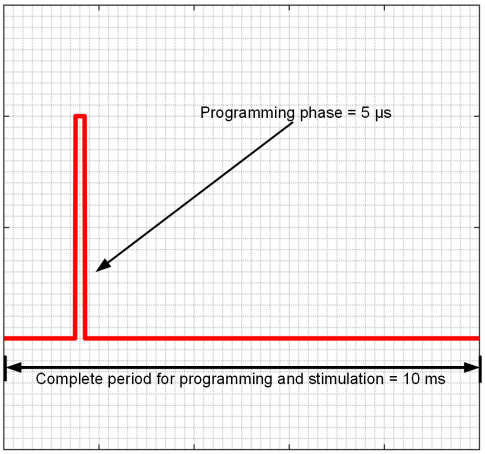


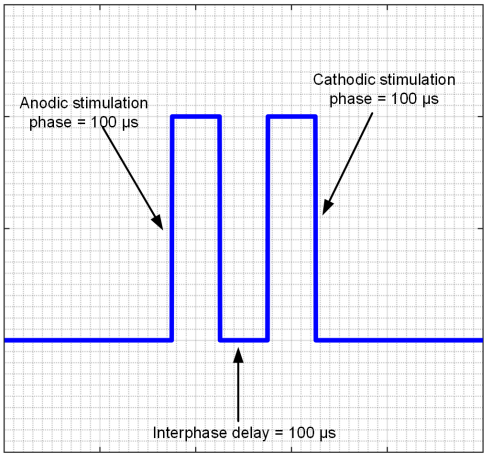
Figure 7.2: Architecture of the DCS interfaced to the binary-weighted anodic and cathodic current sources.

Table 7.2: Timing summary biphasic stimulus.

Phase	Active switches	Clock signals
Programming (φ_1)	$S_1, S_2, S_3, S_4, S_5, S_6$	Programming phase clocks
Stimulation (φ_2)	$S_{N1} \ldots S_{N8}, S_{P1} \ldots S_{P8}, S_{IS1}, S_{IS2}$	Anodic and cathodic stimulation phase clocks



(a)



(b)

Figure 7.3: (a) Programming phase clock. (b) Stimulation phase clock.

switches in Table 7.2 and deactivating all programming phase switches. In terms of bias voltage, this uses a straightforward approach, unlike the current steering system in [128] which needs a linearization circuit for its VCR bias voltage.

7.2.2 Binary Weighted Current Sources

These are activated only during stimulation. From Figure 7.2, transistors M_{P1} to M_{P8} are binary weighted transistors, with the last two, M_{P7} and M_{P8} having the same width. These transistors share the N_{AMP} amplifier and serve as the anodic current sources. While transistors M_{N1} to M_{N8} share the P_{AMP} amplifier and serve as the cathodic current sinks. Both the anodic and cathodic current sources are operated such that all transistors are in saturation with a V_{DSAT} of 100 mV during stimulation. These weighted transistors together with the DCS create 8 dynamic VCR current sources in a push-pull configuration, and are interfaced to the electrodes via a multiplexing stage. The currents through the transistors have a relative weight of $\frac{I}{2}$ for I_{P1} and I_{N1}, $\frac{I}{4}$ for I_{P2} and I_{N2}, and reduces by a factor of 2 up to $\frac{I}{128}$ for I_{P7}, I_{P8}, I_{N7} and I_{N8}, so that the total current sourced or sinked sums up to I, where I is the programming current. This approach has an advantage of increasing the resolution without an increase in current source area, by further splitting the least weighted current source depending on the adopted radix.

7.2.3 Multiplexing Stage

The 8 push-pull current sources are interfaced to the electrodes via a multiplexing stage, as shown in the schematic of Figure 7.4.This stage enables the routing of current to specific electrodes. In order to ensure that each electrode can be individually driven in an anodic or cathodic configuration, the electrodes have independent anode and cathode connections as can be seen in Figure 7.4 labelled A_i and C_i respectively, where i is between 1 and 16. For the 16-electrodes in this design, there are 16 different anodes connected to each of the 8 anodic current sources and 16 different cathodes connected to each of the 8 cathodic current sinks. These connections between the anodes or cathodes to individual current sources is enabled using complementary switching. The complementary switches select which of the weighted current sources to connect to the anodes and cathodes based on control signals. The configuration in which the electrodes operate is determined by the state of the A_i and C_i switches.

7.3 Simulated Results

The output stage was designed in a $0.18\mu m$ CMOS technology and simulated in Cadence Spectre. The clock signals in Table 7.2 and Figure 7.3 were used to simulate the transient analysis of Figure 7.5.

In Figure 7.5, the output stage delivers 503.2 μA and -503.2 μA (programmed for 500 μA) through electrode $E8$ during both anodic and cathodic phases to a $5k\Omega$ load for periods of $100\mu s$. This current is steered to two different electrodes for similar period, to $E1$ delivering $-193.7 : \mu A$ and 189.6 μA (programmed for 187.5 μA) during anodic and cathodic phases and E16 delivering -309.5 μA and 313.6 μA (programmed

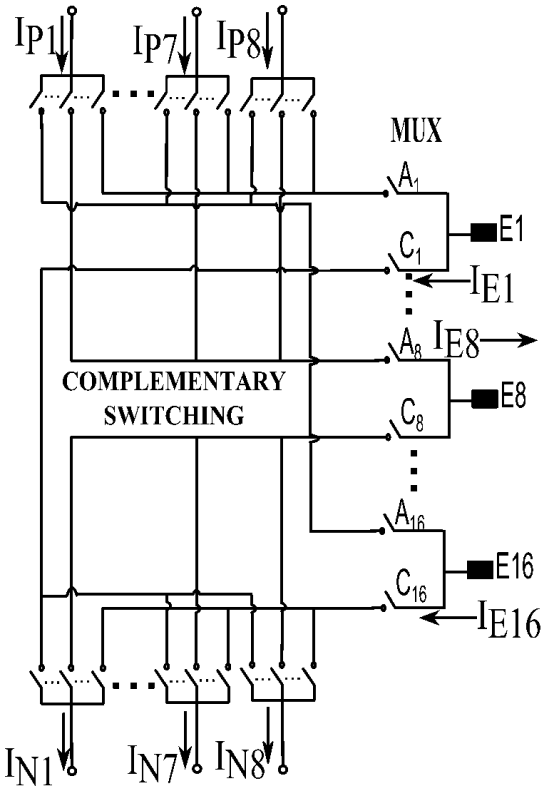


Figure 7.4: Schematic of the proposed multiplexing stage showing how the current sources are interfaced to the stimulating electrodes.

for 312.5 μA) respectively. This is a matching accuracy of within 2%. The DC analysis in Figure 7.6 shows that a headroom voltage of less than 300 mV is needed for the output stage under variable load conditions. In terms of matching, Figure 7.7 compares the input-output characteristics of the output current from the LV and HV DCS.

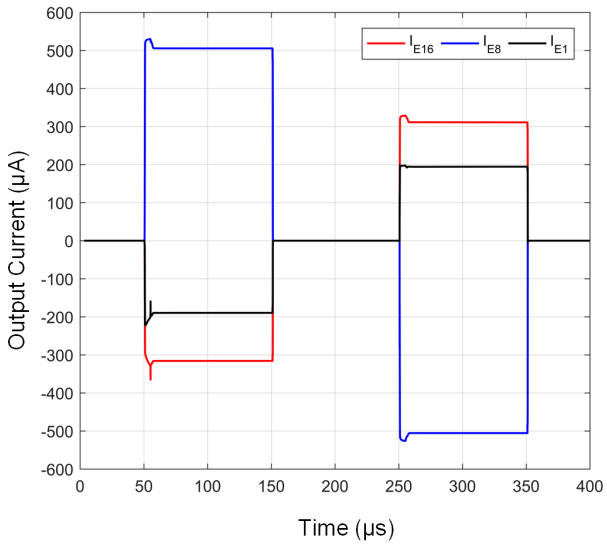


Figure 7.5: Transient operation of current driver (configuration as depicted in Figure 7.4) using $E1$ and $E16$ as cathodes, with $E8$ as anode during initial stimulation phase.

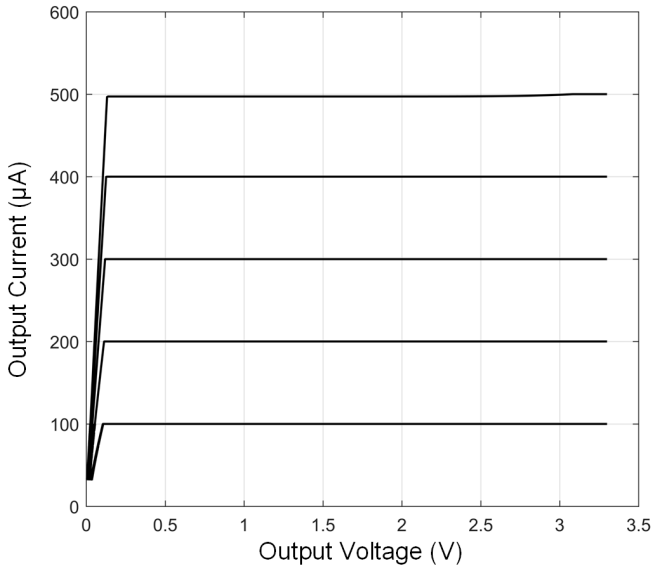


Figure 7.6: Output characteristic depicting the headroom voltage requirements for different levels of output current under load conditions.

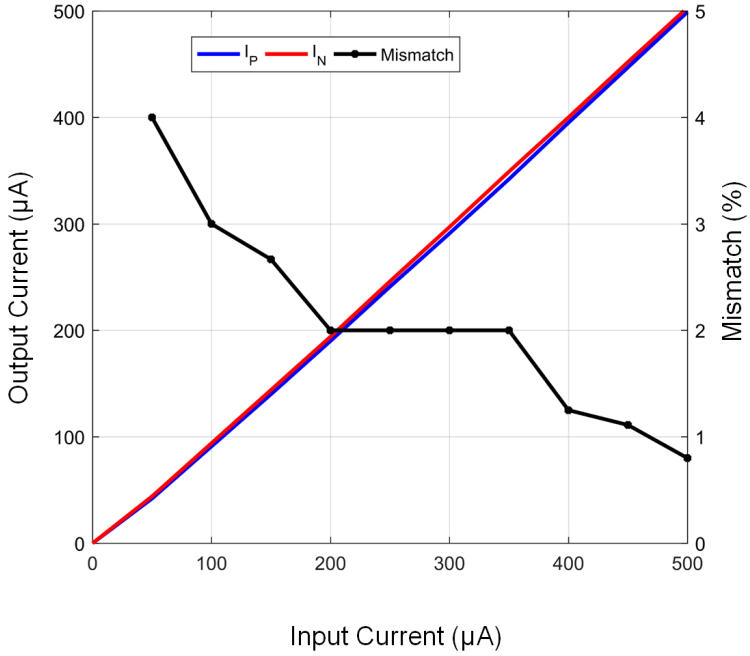


Figure 7.7: Transfer characteristics of the current driver for 5 $k\Omega$ load, I_P (Anodic current) and I_N (Cathodic current).

The mismatch between I_N and I_P is shown to be below 4% throughout the operating range, and can be attributed to a slight difference in bias voltage during programming, which can be improved using larger storage capacitors. However, using larger capacitors increases the area of the circuit which is not desired. Finally, Figure 7.8 depicts the effects of process variations and mismatches by performing 200 runs of Monte Carlo analysis of the output stage for I_{DAC} of 500 μA and 100 μA. The analysis was run

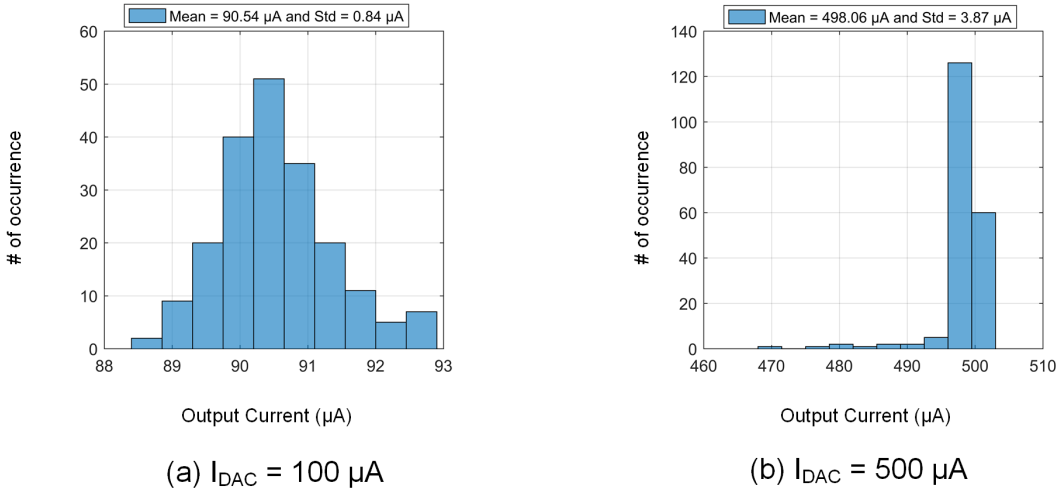


Figure 7.8: Monte Carlo Analysis of effects of process variations and mismatches on the output current I_{OUT} depicted using 10 bins. The analysis was done on 200 runs for $3-\sigma$ models.

using a $3-\sigma$ model with Gaussian distribution on typical corners. For $I_{DAC} = 100\,\mu A$, 77% fall within a 10% tolerance limit, and for $I_{DAC} = 500\,\mu A$, 100% fall within a 10% tolerance limit.

7.4 Discussion

State-of the-art DBS stimulators use a supply voltage greater than 10V. A typical example is the Activa family of stimulators produced by Medtronic which has the ability to administer up to 10.5V, voltage mode stimulation. However, the application of high voltage stimulation is still unused in clinical DBS, as clinical DBS still adopts a maximum stimulation of within 3 - 4 V. Nevertheless, incorporating high voltage capability is the current trend for DBS stimulators marketed by major companies like Boston Scientific, Medtronic and St Jude. A possible explanation for maintaining clinical DBS within low voltages for even the high voltage systems can be attributed to the possibility of brain tissue damage with high voltage stimulations. As an example, consider the conventional Medtronic DBS lead model 3389, which has a diameter of 1.27 mm. And is placed in the STN for stimulation. The STN is a grey matter, having a conductivity of about 0.255 S/m along the height of the cylindrical DBS electrode (with length 1.5mm) [291]. Furthermore, it is also known that the minimum power density that causes brain tissue damage is $800\,\mu W/mm^2$ [39]. By neglecting the impedance as a result of electrode-tissue interface, an approximate value for the maximum voltage the brain tissue around the electrode can withstand is 3.54 V. This is why state-of-the-art DBS systems having high voltage supply, still use a maximum voltage less than 4 V. It is for this reason that the proposed stimulation technique uses a 3.3 V supply as a proof of concept. The aim of the design was to demonstrate the possibility of using a single bias voltage for stimulation during both anodic and cathodic phases of stimulation. This was in order to conserve power and reduce mismatch. Also, the design demonstrated the use of a single current source in producing several weighted current sources that

can be interfaced to a multiplexing stage in order to facilitate current steering. This was in contrast to other directional steering systems like the Vercise current steering system which uses multiple independent current sources [292]. Multiple independent current sources introduce redundancy in the circuit, as not all current sources may be utilised at a given time, this approach promotes inefficiency since it uses an excessive amount of high voltage sources. For the dynamic current steering system in this work, a high voltage alternative will require very large storage capacitors in order to store the bias voltage for use during both stimulation phases. This is the major bottleneck of the design, as this will increase the area of the system. The major factor that will determine the implementation of the high voltage design of the dynamic current steering system is if the increased area as a result of the storage capacitor will serve as a good trade-off for reduced power consumption and mismatch.

7.5 Chapter Summary

a) The aim of the design is to implement an output stage for dynamic current steering which can overcome the common inefficacies associated with DBS, which involve output current mismatch and power consumption.

b) The output stage uses a 3.3 V supply, however it can be scaled for use with higher supply voltages. The output stage lends itself well to increasing current source resolution without a corresponding increase in area or power consumption.

c) The adopted multichannel configuration offers control over current delivered at the stimulation site, which has the potential of improving neural selectivity. Also the use of current controlled stimulation enables constant excitation. And the dynamic current sources assist in reducing power usage and mismatch.

d) With clearer insights into the effects of DBS at the stimulation site, control algorithms can be designed to provide effective stimulation paradigms that can be used in implementing dynamic current steering. Future developments will incorporate novel charge balancing techniques.

Chapter 8

Conclusion and Future Directions

8.1 Summary

Clinical DBS is a tool used to mitigate pharmacologically intractable neurological disorders such as PD, ET and dystonia. It currently uses continuous, high frequency voltage or current pulses to mitigate PD. This results in a number of limitations, notable among are stimulation induced side effects and shortening of pacemaker battery life. These limitations can be tackled using aDBS, which delivers stimulation precisely only when needed. This thesis presents work undertaken to investigate, propose and develop novel algorithms and implementations for adapting DBS. The major focus of the work was to identify implementation strategies that could lead to fully implantable systems for aDBS, vis-a-vis biomarker processing and stimulator implementation. Biomarker processing involves strategies to detect PD such that PD detection informs therapy. Thus, enabling the use of demand-driven DBS to overcome the shortcomings of continuous DBS. For stimulator implementation, a directional steering stimulator with capabilities for spatially adaptating DBS was proposed. Spatial adaptation offers a new dimension to DBS therapy by directional control over stimulation focus. These two forms of DBS adaptation motivated the major contributions of the work.

To enable hardware-efficient PD diagnosis and management, the first part of the work involved establishing a processing chain for PD detection and evaluating a combination of algorithms across the processing chain based on customised performance metrics. Selection of appropriate performance metrics are the cornerstone for a thorough assessment of PD detection algorithms. This motivated the use of custom as well as standard measures for evaluating the algorithms in terms of area, NOP and efficacy. Establishing a detection scheme consisting of dynamic FE and dynamic pattern classification was the first novelty of the work. This was tagged dynamic detection. The PD detection algorithms were designed to incorporate dynamic detection schemes such that feature extraction and classification were made dynamic so as to track the unpredictable nature of Parkinsonian LFP. The need for dynamic detection was essential due to the non-linear progression of PD in patients. Furthermore, a novel DR technique, the MRM (inspired by Fisher's separability criterion) was proposed as part of the algorithms to be evaluated. It provided the most efficient DR performance; based on customised

complexity and accuracy metrics. The PD detection algorithms were validated using semi-synthetic LFP data generated from representative PD and non-PD LFP recordings. The statistical methods used in LFP data synthesis was another novelty of the work.

For hardware demonstration, a processor consisting of the combination with the best accuracy and complexity trade-offs was implemented and prototyped on an FPGA platform. This was the first known hardware implementation of a neurological event detector for PD. To establish the most optimal implementation of the detector, various implementations of its classification stage were analysed for efficacy, power and area trade-offs. More emphasis was placed on optimising the classifier as it is the most computationally intensive processing stage. The implementation was proposed for use in personalised PD monitoring. Personalised health monitoring can serve as a precursor to fully implantable aDBS systems. The major novelty in the PD processor was the choice of application, as it is the first known hardware implementation intended for on-site PD detection and monitoring. Another novelty involved the detection approach, which uses complementary detection. This involved using a number of weak classifiers to produce a classifier with improved consistency and confidence level. Since more knowledge regarding PD and DBS is still gained by the day, a microchip implementation of the design may not be cost effective at this point. This prompted prototyping it on an FPGA platform because RTL implementations could be seamlessly translated for use on microcontrollers or implemented as a microchip. Nevertheless, microchip performance analysis for power and area were obtained to guide future implementations. This analysis can serve as a guide for anyone embarking on ASIC implementation of PD detection processors. The PD detection processor was aimed for use in monitoring PD onset such that the information is communicated for external use by caregivers or as a trigger for stimulation devices; both of which could be used to mitigate patients conditions.

The thesis also proposes a smart control strategy for adapting DBS that is designed to overcome the high efficiency demands of aDBS without compromising on hardware resources. It proposes a critic-actor control approach that uses a LUT based controller implemented using a fuzzy logic controller. The critic-actor approach was adopted because it models the relationship between the physician and the automated neuromodulation system. Like the trained clinician, the critic evaluates the patient based on a defined criteria. The actor "acts" based on information from the critic. The motivation for adopting this approach was inspired by the need to incorporate more heuristics in PD suppression due to the extreme difficulty in taming PD as well as the limited understanding of its internal mechanisms. Besides, the requirement that hardware implementations of aDBS should be designed to take into consideration implantability also motivated the critic-actor approach. Nevertheless, further analysis may be needed to assess the effect of external body signals as well as an increased number of input signals on the performance of the controller.

With the first three parts focused on PD detection and tracking for aDBS, the fourth

and final part presents preliminary design on a dynamic current steering stimulator output stage that lends itself to spatial adaptation. A full implementation of the design could be used for controlling activation field spread beyond targeted areas, which is a major source of stimulation induced side effects.

8.2 Original Contributions

The work in this thesis provides the following significant contributions:

1. Chapter 4 proposed dynamic PD detection as a way to overcome the inconsistencies associated with neural signals during PD detection. This led to the development of an algorithm for feature selection, inspired by Fisher's separability criterion: the MRM. The MRM implemented a hardware-efficient feature and channel selection of LFP power features.

2. Chapter 4 also embarked on the development of the first known methodology for a balanced, objective and fair evaluation of trade-offs between accuracy and computational complexity for PD detection algorithms using customised performance metrics. A subset of algorithms consisting of FE, DR and dynamic pattern classification stages were analysed and investigated for hardware-efficiency.

3. Section 4.2 in Chapter 4 presented the first known statistical method for semi-synthetic LFP generation that can be used for unbiased and accurate evaluation of PD detection algorithms.

4. Chapter 5 was involved with the development of the first known hardware implementation of a processor for PD detection and monitoring. Based on power, area and efficacy trade-offs, complementary detection was presented as a consistent and hardware-efficient method of PD detection and monitoring.

5. Chapter 6 proposed a framework for adapting DBS using a critic-actor control strategy. The approach models the relationship between the physician and the automated neuromodulation system. The critic like the "trained clinician" assesses the state of the system based on a cost function (in this case state estimates) and provides the information to the actor.

6. Chapter 7 introduces the design of an output stage for dynamic current steering which is aimed towards the development of a power-efficient multi-polar multi-site stimulator for spatially adapting DBS.

8.3 Suggestions for Future Work

This section discusses some ways the work presented in this thesis can be developed further.

8.3.1 Implementation of a Power-Efficient Output Stage for Multisite and Multipolar DBS

Therapeutic benefits of DBS are often affected by poor stimulation focus [99]. This is one of the major motivations for multisite and multipolar DBS. In addition, precise targeting during stimulation has been reported to reduce the threshold for beneficial effect [125]. This makes the design of power efficient stimulator output stages essential. Boston Scientific, Medtronic and Aleva Neuroprosthetics are currently marketing directional steering stimulators. However, their designs use multiple current sources, which is an inefficient approach. The need for a power-efficient approach motivated the work in Chapter 7 which introduced a possible implementation of a dynamic current steering stimulator. The design is a preliminary implementation. For future implementations, the design can be extended to use a large number of electrodes. Also, the design specification can be further improved by increasing the output current to 20 mA. Above all, the output stage could be developed into a full ASIC.

8.3.2 Development of a Hardware Efficient PD Prediction Processor

Machine learning provides techniques for analysing complex physiological signals, which was why it was used for PD tracking in this work. PD tracking presented in this work and in other aDBS applications have mainly focussed on early detection and state estimation. However, for optimal therapy, prediction of critical events may be more beneficial than early detection. The major challenge of prediction is that critical events preceding PD may be difficult to discern due to the unpredictable nature of PD dynamics. Nevertheless, it will be important in the future to explore the possibility of predicting PD events in advance. A starting point in PD prediction could be the adoption of strategies that combine short-term (like electrical activity measurements) and long-term (like metabolic activity measurements) sensing such that repetitive patterns from the long term sensors can be used to assist in PD prediction. A major issue in achieving this will be the need to design novel DBS recording electrodes that can concurrently obtain electrical and metabolic activity measurements from the brain. For implantability, it is imperative that the prediction processor or system is designed with hardware implementation in mind.

8.3.3 Investigation into the Efficacy and Hardware Efficiency of Closed-loop Control Strategies

For closed-loop DBS, clever control strategies are required to generate actionable outputs. However, for fully implantable systems, there are certain hardware resource constraints imposed. It will thus be important to embark on a study that works towards answering the following questions:

- Which control strategies are most effective for modulating therapy?
- Which control strategies are most robust to noise and other electrical or metabolic disturbance sources in patients?

- Which control strategies are the most feasible for implantable hardware implementation?
- How sophisticated do the control techniques need to be and how much can they adapt to effects like disease progression, environmental factors, mechanical factors, and behaviourally induced changes in brain or network activity?

Due to the complexity and limited understanding of the brain, efficacious and hardware efficient controller design for aDBS could be the next frontier of DBS research.

References

[1] M. L. Kringelbach, N. Jenkinson, S. L. F. Owen, and T. Z. Aziz, "Translational principles of deep brain stimulation." *Nature reviews. Neuroscience*, vol. 8, no. 8, pp. 623–635, 2007.

[2] "Basal Ganglia — Upright-Health.com." [Online]. Available: http://www.upright-health.com/basal-ganglia/

[3] S. J. Schiff, *Neural Control Engineering: The Emerging Intersection Between Control Theory and Neuroscience*. Cambridge, Massachusetts: The MIT Press, 2012.

[4] Medgadget, "Vercise Deep Brain Neurostimulator EU Approved for Tremor," 2014. [Online]. Available: http://www.medgadget.com/2014/09/vercise-deep-brain-neurostimulator-eu-approved-for-tremor-video.html

[5] A. Espay, "Parkinson Surgery," 2009. [Online]. Available: https://www.mayfieldclinic.com/PE-SurgPD.htm

[6] M. Sadilek and S. Thurner, "Physiologically motivated multiplex Kuramoto model describes phase diagram of cortical activity," *Scientific Reports*, vol. 5, p. 10015, may 2015.

[7] D. M. Durand, "Electrical Stimulation of Excitable Tissue," in *The Biomedical Engineering Handbook*, 2nd ed., J. D. Bronzino, Ed. Boca Raton, Florida: CRC Press, 2000, p. 2896.

[8] H. C. F. Martens, E. Toader, M. M. J. Decré, D. J. Anderson, R. Vetter, D. R. Kipke, K. B. Baker, M. D. Johnson, and J. L. Vitek, "Spatial steering of deep brain stimulation volumes using a novel lead design," *Clinical Neurophysiology*, vol. 122, no. 3, pp. 558–566, 2011.

[9] A. O. Hebb, J. J. Zhang, M. H. Mahoor, C. Tsiokos, C. Matlack, H. J. Chizeck, and N. Pouratian, "Creating the Feedback Loop: closed-loop neurostimulation," *Neurosurgery Clinics of North America*, vol. 25, no. 1, pp. 187–204, jan 2014.

[10] S. Marshland, *Machine Learning : an algorithmic perspective*, 2nd ed. Boca Raton, FL: CRC Press, 2015.

[11] A. L. Benabid, P. Pollak, C. Gervason, D. Hoffmann, D. M. Gao, M. Hommel, J. E. Perret, and J. de Rougemont, "Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus." *Lancet (London, England)*, vol. 337, no. 8738, pp. 403–6, feb 1991.

[12] H. E. Ward, N. Hwynn, and M. S. Okun, "Update on deep brain stimulation for neuropsychiatric disorders." *Neurobiology of disease*, vol. 38, no. 3, pp. 346–53, jun 2010.

[13] N. Choices, "Parkinson's disease - NHS Choices," 2016. [Online]. Available: http://www.nhs.uk/conditions/parkinsons-disease/Pages/Introduction.aspx

[14] J. M. Beitz, "Parkinson's disease: a review." *Frontiers in bioscience (Scholar edition)*, vol. 6, pp. 65–74, jan 2014.

[15] W. Dauer and S. Przedborski, "Parkinson's disease: mechanisms and models." *Neuron*, vol. 39, no. 6, pp. 889–909, sep 2003.

[16] S. Little and P. Brown, "What brain signals are suitable for feedback control of deep brain stimulation in Parkinson's disease?" *Annals of the New York Academy of Sciences*, vol. 1265, no. 1, pp. 9–24, 2012.

[17] K. A. Follett, F. M. Weaver, M. Stern, K. Hur, C. L. Harris, P. Luo, W. J. Marks, J. Rothlind, O. Sagher, C. Moy, R. Pahwa, K. Burchiel, P. Hogarth, E. C. Lai, J. E. Duda, K. Holloway, A. Samii, S. Horn, J. M. Bronstein, G. Stoner, P. A. Starr, R. Simpson, G. Baltuch, A. De Salles, G. D. Huang, and D. J. Reda, "Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease." *The New England journal of medicine*, vol. 362, no. 22, pp. 2077–91, jun 2010.

[18] "Surgical treatment for Parkinson's disease." *The Lancet. Neurology*, vol. 3, no. 12, pp. 719–28, dec 2004.

[19] E. B. Montgomery and J. T. Gale, "Mechanisms of action of deep brain stimulation(DBS) ." *Neuroscience and biobehavioral reviews*, vol. 32, no. 3, pp. 388–407, jan 2008.

[20] S. Miocinovic, S. Somayajula, S. Chitnis, and J. L. Vitek, "History, applications, and mechanisms of deep brain stimulation." *JAMA neurology*, vol. 70, no. 2, pp. 163–71, feb 2013.

[21] R. Carron, A. Chaillet, A. Filipchuk, W. Pasillas-Lépine, and C. Hammond, "Closing the loop of deep brain stimulation." *Frontiers in systems neuroscience*, vol. 7, no. December, p. 112, 2013.

[22] J. Volkmann, J. Herzog, F. Kopper, and G. Deuschl, "Introduction to the programming of deep brain stimulators." *Movement disorders : official journal of the Movement Disorder Society*, vol. 17 Suppl 3, pp. S181–7, jan 2002.

[23] P. Hickey and M. Stacy, "Deep Brain Stimulation: A Paradigm Shifting Approach to Treat Parkinson's Disease," *Frontiers in Neuroscience*, vol. 10, p. 173, apr 2016.

[24] S. Little, A. Pogosyan, S. Neal, B. Zavala, L. Zrinzo, M. Hariz, T. Foltynie, P. Limousin, K. Ashkan, J. Fitzgerald, A. L. Green, T. Z. Aziz, and P. Brown, "Adaptive deep brain stimulation in advanced Parkinson disease," *Annals of Neurology*, pp. 449–457, 2013.

[25] A. C. Meidahl, G. Tinkhauser, D. M. Herz, H. Cagnan, J. Debarros, and P. Brown, "Adaptive Deep Brain Stimulation for Movement Disorders: The Long Road to Clinical Therapy," *Movement Disorders*, vol. 32, no. 6, pp. 810–819, jun 2017.

[26] J. Volkmann, "Deep brain stimulation for the treatment of Parkinson's disease." *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 21, no. 1, pp. 6–17, jan 2004.

[27] C. W. Hess, D. E. Vaillancourt, and M. S. Okun, "The temporal pattern of stimulation may be important to the mechanism of deep brain stimulation," *Experimental Neurology*, vol. 247, pp. 296–302, sep 2013.

[28] A. L. Benabid, P. Pollak, C. Gross, D. Hoffmann, A. Benazzouz, D. M. Gao, A. Laurent, M. Gentil, and J. Perret, "Acute and long-term effects of subthalamic nucleus stimulation in Parkinson's disease." *Stereotactic and functional neurosurgery*, vol. 62, no. 1-4, pp. 76–84, 1994.

[29] A. Benazzouz, C. Gross, J. Féger, T. Boraud, and B. Bioulac, "Reversal of rigidity and improvement in motor performance by subthalamic high-frequency stimulation in MPTP-treated monkeys." *The European journal of neuroscience*, vol. 5, no. 4, pp. 382–9, apr 1993.

[30] M. Lafreniere-Roula, E. Kim, W. D. Hutchison, A. M. Lozano, M. Hodaie, and J. O. Dostrovsky, "High-frequency microstimulation in human globus pallidus and substantia nigra," *Experimental Brain Research*, vol. 205, no. 2, pp. 251–261, aug 2010.

[31] T. Boraud, E. Bezard, B. Bioulac, and C. Gross, "High frequency stimulation of the internal Globus Pallidus (GPi) simultaneously improves parkinsonian symptoms and reduces the firing frequency of GPi neurons in the MPTP-treated monkey." *Neuroscience letters*, vol. 215, no. 1, pp. 17–20, aug 1996.

[32] "Mechanism of Deep Brain Stimulation: Inhibition, Excitation, or Disruption?" *The Neuroscientist*, vol. 22, no. 3, pp. 313–322, jun 2016.

[33] J. D. Carlson, D. R. Cleary, J. S. Cetas, M. M. Heinricher, and K. J. Burchiel, "Deep Brain Stimulation Does Not Silence Neurons in Subthalamic Nucleus in Parkinson's Patients," *Journal of Neurophysiology*, vol. 103, no. 2, pp. 962–967, feb 2010.

[34] T. M. Herrington, J. J. Cheng, and E. N. Eskandar, "Mechanisms of deep brain stimulation," *Journal of Neurophysiology*, vol. 115, no. 1, pp. 19–38, jan 2016.

[35] A. Kupsch, M. Tagliati, M. Vidailhet, T. Aziz, P. Krack, E. Moro, and J. K. Krauss, "Early postoperative management of DBS in dystonia: programming, response to stimulation, adverse events, medication changes, evaluations, and troubleshooting." *Movement disorders : official journal of the Movement Disorder Society*, vol. 26 Suppl 1, pp. S37–53, jun 2011.

[36] J. F. Baizabal-Carvallo and J. Jankovic, "Movement disorders induced by deep brain stimulation." *Parkinsonism & related disorders*, jan 2016.

[37] M. L. Hacker, J. Tonascia, M. Turchan, A. Currie, L. Heusinkveld, P. E. Konrad, T. L. Davis, J. S. Neimat, F. T. Phibbs, P. Hedera, L. Wang, Y. Shi, D. M. Shade, A. L. Sternberg, L. T. Drye, and D. Charles, "Deep brain stimulation may reduce the relative risk of clinically important worsening in early stage Parkinson's disease." *Parkinsonism & related disorders*, vol. 21, no. 10, pp. 1177–83, oct 2015.

[38] M. Arlotti, M. Rosa, S. Marceglia, S. Barbieri, and A. Priori, "The adaptive deep brain stimulation challenge." *Parkinsonism & related disorders*, vol. 28, pp. 12–7, jul 2016.

[39] T. M. Seese, H. Harasaki, G. M. Saidel, and C. R. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of

muscle tissue to chronic heating." *Laboratory investigation; a journal of technical methods and pathology*, vol. 78, no. 12, pp. 1553–62, dec 1998.

[40] M. R. DeLong and T. Wichmann, "Circuits and Circuit Disorders of the Basal Ganglia," *Archives of Neurology*, vol. 64, no. 1, p. 20, jan 2007.

[41] M. Hariz, P. Blomstedt, and L. Zrinzo, "Future of brain stimulation: New targets, new indications, new technology," *Movement Disorders*, vol. 28, no. 13, pp. 1784–1792, nov 2013.

[42] J. W. Mink, "The basal ganglia: focused selection and inhibition of competing motor programs." *Progress in neurobiology*, vol. 50, no. 4, pp. 381–425, nov 1996.

[43] J. C. Giugni and M. S. Okun, "Treatment of advanced Parkinson's disease." *Current opinion in neurology*, vol. 27, no. 4, pp. 450–60, aug 2014.

[44] C. O. Oluigbo, A. Salma, and A. R. Rezai, "Deep Brain Stimulation for Neurological Disorders," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 88–99, 2012.

[45] A. Amon and F. Alesch, "Systems for deep brain stimulation: review of technical features," *Journal of Neural Transmission*, pp. 1–9, jul 2017.

[46] Medtronic, "Medtronic DBS Therapy for Parkinsons Disease and Essential Tremor Clinical Summary," 2017. [Online]. Available: http://www.medtronic.com/us-en/healthcare-professionals/therapies-procedures/neurological/deep-brain-stimulation.html

[47] M. S. Okun and P. R. Zeilman, "Parkinson's Disease: Guide to Deep Brain Stimulation," National Parkinson Foundation, Florida, Tech. Rep., 2017. [Online]. Available: http://www.parkinson.org/sites/default/files/Guide{_}to{_}DBS{_}Stimulation{_}Therapy.pdf

[48] Boston Scientific, "Deep Brain Stimulation Systems — Boston Scientific," 2017. [Online]. Available: http://www.bostonscientific.com/en-EU/products/deep-brain-stimulation-systems.html

[49] St Jude Medical, "Deep Brain Stimulation IPG Device — St. Jude Medical," 2017. [Online]. Available: https://www.sjmglobal.com/en-int/professionals/featured-products/neuromodulation/deep-brain-stimulation/implantable-pulse-generators

[50] A. L. HODGKIN and A. F. HUXLEY, "A quantitative description of membrane current and its application to conduction and excitation in nerve." *The Journal of physiology*, vol. 117, no. 4, pp. 500–44, aug 1952.

[51] R. Fitzhugh, "Impulses and Physiological States in Theoretical Models of Nerve Membrane." *Biophysical journal*, vol. 1, no. 6, pp. 445–66, jul 1961.

[52] H. R. Wilson and J. D. Cowan, "Excitatory and inhibitory interactions in localized populations of model neurons." *Biophysical journal*, vol. 12, no. 1, pp. 1–24, jan 1972.

[53] J. Ziburkus, J. R. Cressman, and S. J. Schiff, "Seizures as imbalanced up states: excitatory and inhibitory conductances during seizure-like events." *Journal of neurophysiology*, vol. 109, no. 5, pp. 1296–306, mar 2013.

[54] Y. Shu, A. Hasenstaub, and D. A. McCormick, "Turning on and off recurrent balanced cortical activity," *Nature*, vol. 423, no. 6937, pp. 288–293, may 2003.

[55] B. Haider, A. Duque, A. R. Hasenstaub, and D. A. McCormick, "Neocortical Network Activity In Vivo Is Generated through a Dynamic Balance of Excitation and Inhibition," *Journal of Neuroscience*, vol. 26, no. 17, pp. 4535–4545, apr 2006.

[56] J. D. Cowan, J. Neuman, and W. van Drongelen, "Wilson-Cowan Equations for Neocortical Dynamics." *Journal of mathematical neuroscience*, vol. 6, no. 1, p. 1, dec 2016.

[57] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia." *Nature*, vol. 442, no. 7099, pp. 164–71, jul 2006.

[58] P. Afshar, A. Khambhati, S. Stanslaski, D. Carlson, R. Jensen, D. Linde, S. Dani, M. Lazarewicz, P. Cong, J. Giftakis, P. Stypulkowski, and T. Denison, "A translational platform for prototyping closed-loop neuromodulation systems." *Frontiers in neural circuits*, vol. 6, no. January, p. 117, 2012.

[59] J. D. Rolston, R. E. Gross, and S. M. Potter, "Closed-loop, open-source electrophysiology." *Frontiers in neuroscience*, vol. 4, jan 2010.

[60] A. A. Kuhn, A. Kupsch, G.-H. Schneider, and P. Brown, "Reduction in subthalamic 8-35 Hz oscillatory activity correlates with clinical improvement in Parkinson's disease." *The European journal of neuroscience*, vol. 23, no. 7, pp. 1956–60, apr 2006.

[61] B. Lysyansky, O. V. Popovych, and P. A. Tass, "Desynchronizing anti-resonance effect of m: n ON-OFF coordinated reset stimulation." *Journal of neural engineering*, vol. 8, no. 3, p. 036019, jun 2011.

[62] M. Weinberger, N. Mahant, W. D. Hutchison, A. M. Lozano, E. Moro, M. Hodaie, A. E. Lang, and J. O. Dostrovsky, "Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in Parkinson's disease." *Journal of neurophysiology*, vol. 96, no. 6, pp. 3248–56, dec 2006.

[63] S. Y. Kung, Y. Luo, and M.-W. Mak, "Feature Selection for Genomic Signal Processing: Unsupervised, Supervised, and Self-Supervised Scenarios," *Journal of Signal Processing Systems*, vol. 61, no. 1, pp. 3–20, oct 2008.

[64] K. H. Lee, S. Y. Kung, and N. Verma, "Improving kernel-energy trade-offs for machine learning in implantable and wearable biomedical applications," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 1597–1600, 2011.

[65] G. Meyfroidt, F. Güiza, J. Ramon, and M. Bruynooghe, "Machine learning techniques to examine large patient databases." *Best practice & research. Clinical anaesthesiology*, vol. 23, no. 1, pp. 127–43, mar 2009.

[66] J. Predd, S. Kulkarni, and H. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, jul 2006.

[67] X.-J. Feng, B. Greenwald, H. Rabitz, E. Shea-Brown, and R. Kosut, "Toward closed-loop optimization of deep brain stimulation for Parkinson's disease: concepts and lessons from a computational model." *Journal of neural engineering*, vol. 4, no. 2, pp. L14–L21, 2007.

[68] P. F. Grant and M. M. Lowery, "Simulation of basal ganglia parkinsonian oscillations and their suppression by closed loop deep brain stimulation," *In Review*, vol. 21, no. 4, pp. 584–594, 2011.

[69] S. Santaniello, G. Fiengo, L. Glielmo, and W. M. Grill, "Closed-loop control of deep brain stimulation: a simulation study." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 19, no. 1, pp. 15–24, 2011.

[70] S. F. Lempka, M. D. Johnson, M. A. Moffitt, K. J. Otto, D. R. Kipke, and C. C. McIntyre, "Theoretical analysis of intracortical microelectrode recordings." *Journal of neural engineering*, vol. 8, no. 4, p. 045006, aug 2011.

[71] E. Bakstein, J. Burgess, K. Warwick, V. Ruiz, T. Aziz, and J. Stein, "Parkinsonian tremor identification with multiple local field potential feature classification," *Journal of Neuroscience Methods*, vol. 209, no. 2, pp. 320–330, 2012.

[72] G. Giannicola, M. Rosa, D. Servello, C. Menghetti, G. Carrabba, C. Pacchetti, R. Zangaglia, F. Cogiamanian, E. Scelzo, S. Marceglia, L. Rossi, and A. Priori, "Subthalamic local field potentials after seven-year deep brain stimulation in Parkinson's disease," *Experimental Neurology*, vol. 237, no. 2, pp. 312–317, 2012.

[73] U. Rutishauser, E. M. Schuman, and A. N. Mamelak, "Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo." *Journal of neuroscience methods*, vol. 154, no. 1-2, pp. 204–24, jun 2006.

[74] P. J. Grahn, G. W. Mallory, O. U. Khurram, B. M. Berry, J. T. Hachmann, A. J. Bieber, K. E. Bennet, H. K. Min, S. Y. Chang, K. H. Lee, and J. L. Lujan, "A neurochemical closed-loop controller for deep brain stimulation: Toward individualized smart neuromodulation therapies," *Frontiers in Neuroscience*, vol. 8, no. 8 JUN, pp. 1–11, 2014.

[75] B. Rosin, M. Slovik, R. Mitelman, M. Rivlin-Etzion, S. N. Haber, Z. Israel, E. Vaadia, and H. Bergman, "Closed-loop deep brain stimulation is superior in ameliorating parkinsonism," *Neuron*, vol. 72, no. 2, pp. 370–384, 2011.

[76] J.-S. Brittain, P. Probert-Smith, T. Aziz, and P. Brown, "Tremor Suppression by Rhythmic Transcranial Current Stimulation," *Current Biology*, vol. 23, no. 5, pp. 436–440, mar 2013.

[77] H. Cagnan, D. Pedrosa, S. Little, A. Pogosyan, B. Cheeran, T. Aziz, A. Green, J. Fitzgerald, T. Foltynie, P. Limousin, L. Zrinzo, M. Hariz, K. J. Friston, T. Denison, and P. Brown, "Stimulating at the right time: phase-specific deep brain stimulation," *Brain*, vol. 140, no. 1, pp. 132–145, jan 2017.

[78] S. Little, M. Beudel, L. Zrinzo, T. Foltynie, P. Limousin, M. Hariz, S. Neal, B. Cheeran, H. Cagnan, J. Gratwicke, T. Z. Aziz, A. Pogosyan, and P. Brown, "Bilateral adaptive deep brain stimulation is effective in Parkinson's disease." *Journal of neurology, neurosurgery, and psychiatry*, pp. jnnp–2015–310 972–, sep 2015.

[79] L. A. Johnson, S. D. Nebeck, A. Muralidharan, M. D. Johnson, K. B. Baker, and J. L. Vitek, "Closed-Loop Deep Brain Stimulation Effects on Parkinsonian Motor Symptoms in a Non-Human Primate Is Beta Enough?" 2016.

[80] M. G. Rosenblum and A. S. Pikovsky, "Controlling synchronization in an ensemble of globally coupled oscillators." *Physical review letters*, vol. 92, no. 11, p. 114102, mar 2004.

[81] O. Omelchenko, C. Hauptmann, Y. Maistrenko, and P. Tass, "Collective dynamics of globally coupled phase oscillators under multisite delayed feedback stimulation," *Physica D: Nonlinear Phenomena*, vol. 237, no. 3, pp. 365–384, mar 2008.

[82] P. F. Grant and M. M. Lowery, "Simulation of cortico-basal ganglia oscillations and their suppression by closed loop deep brain stimulation." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 21, no. 4, pp. 584–94, jul 2013.

[83] C. Liu, J. Wang, H. Li, M. Lu, B. Deng, H. Yu, X. Wei, C. Fietkiewicz, and K. A. Loparo, "Closed-Loop Modulation of the Pathological Disorders of the Basal Ganglia Network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 2, pp. 371–382, feb 2017.

[84] A. B. Schwartz, "CORTICAL NEURAL PROSTHETICS," *Annual Review of Neuroscience*, vol. 27, no. 1, pp. 487–507, jul 2004.

[85] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc., 2001.

[86] M. J. Morrell and RNS System in Epilepsy Study Group, "Responsive cortical stimulation for the treatment of medically intractable partial epilepsy," *Neurology*, vol. 77, no. 13, pp. 1295–1304, sep 2011.

[87] C. Camara, P. Isasi, K. Warwick, V. Ruiz, T. Aziz, J. Stein, and E. Bak, "Resting tremor classification and detection in Parkinson s disease patients," *Biomedical Signal Processing and Control*, vol. 16, pp. 88–97, 2015.

[88] K. Kostoglou, K. P. Michmizos, P. Stathis, D. Sakas, K. S. Nikita, and G. D. Mitsis, "Classification and Prediction of Clinical Improvement in Deep Brain Stimulation from Intraoperative Microelectrode Recordings," *IEEE Transactions on Biomedical Engineering*, pp. 1–1, jul 2016.

[89] I. Basu, D. Graupe, D. Tuninetti, P. Shukla, K. V. Slavin, L. V. Metman, and D. M. Corcos, "Pathological tremor prediction using surface electromyogram and acceleration: potential use in 'ON-OFF' demand driven deep brain stimulator design." *Journal of neural engineering*, vol. 10, no. 3, p. 036019, jun 2013.

[90] A. Mohammed, M. Zamani, R. Bayford, and A. Demosthenous, "Toward On-Demand Deep Brain Stimulation Using Online Parkinsons Disease Prediction Driven by Dynamic Detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2441–2452, dec 2017.

[91] ——, "Patient specific Parkinson's disease detection for adaptive deep brain stimulation," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, aug 2015, pp. 1528–1531.

[92] F. Rattay, "The basic mechanism for the electrical stimulation of the nervous system." *Neuroscience*, vol. 89, no. 2, pp. 335–46, mar 1999.

[93] D. R. Merrill, M. Bikson, and J. G. R. Jefferys, "Electrical stimulation of excitable tissue: design of efficacious and safe protocols." *Journal of neuroscience methods*, vol. 141, no. 2, pp. 171–98, feb 2005.

[94] X. F. Wei and W. M. Grill, "Impedance characteristics of deep brain stimulation electrodes in vitro and in vivo." *Journal of neural engineering*, vol. 6, no. 4, p. 046008, aug 2009.

[95] A.-L. Benabid, B. Wallace, J. Mitrofanis, C. Xia, B. Piallat, V. Fraix, A. Batir, P. Krack, P. Pollak, and F. Berger, "Therapeutic electrical stimulation of the central nervous system." *Comptes rendus biologies*, vol. 328, no. 2, pp. 177–86, feb 2005.

[96] J. Simpson and M. Ghovanloo, "An Experimental Study of Voltage, Current, and Charge Controlled Stimulation Front-End Circuitry," *2007 IEEE International Symposium on Circuits and Systems*, pp. 325–328, 2007.

[97] J. M. Bronstein, M. Tagliati, C. McIntyre, R. Chen, T. Cheung, E. L. Hargreaves, Z. Israel, M. Moffitt, E. B. Montgomery, P. Stypulkowski, J. Shils, T. Denison, J. Vitek, J. Volkman, J. Wertheimer, and M. S. Okun, "The Rationale Driving the Evolution of Deep Brain Stimulation to Constant-Current Devices," *Neuromodulation: Technology at the Neural Interface*, vol. 18, no. 2, pp. 85–89, feb 2015.

[98] A. Wagle Shukla and M. S. Okun, "Surgical Treatment of Parkinsons Disease: Patients, Targets, Devices, and Approaches," *Neurotherapeutics*, vol. 11, no. 1, pp. 47–59, jan 2014.

[99] A. Lavano, A. D. Torre, and G. Guzzi, "Directional Deep Brain Stimulation," *Brain Disorders & Therapy*, vol. 05, no. 03, pp. 1–3, sep 2016.

[100] "Proceedings of the Third Annual Deep Brain Stimulation Think Tank: A Review of Emerging Issues and Technologies," *Frontiers in Neuroscience*, vol. 10, p. 119, apr 2016.

[101] C. C. McIntyre, A. Chaturvedi, R. R. Shamir, and S. F. Lempka, "Engineering the Next Generation of Clinical Deep Brain Stimulation Technology," *Brain Stimulation*, vol. 8, no. 1, pp. 21–26, jan 2015.

[102] E. W. Tsang, C. Hamani, E. Moro, F. Mazzella, U. Saha, A. M. Lozano, M. Hodaie, R. Chuang, T. Steeves, S. Y. Lim, B. Neagu, and R. Chen, "Subthalamic deep brain stimulation at individualized frequencies for Parkinson disease," *Neurology*, vol. 78, no. 24, pp. 1930–1938, jun 2012.

[103] S.-Y. Chang, C. J. Kimble, I. Kim, S. B. Paek, K. R. Kressin, J. B. Boesche, S. V. Whitlock, D. R. Eaker, A. Kasasbeh, A. E. Horne, C. D. Blaha, K. E. Bennet, and K. H. Lee, "Development of the Mayo Investigational Neuromodulation Control System: toward a closed-loop electrochemical feedback system for deep brain stimulation," *Journal of Neurosurgery*, vol. 119, no. 6, pp. 1556–1565, dec 2013.

[104] M. Rosenblum and A. Pikovsky, "Delayed feedback control of collective synchrony:An approach to suppression of pathological brain rhythms," *Physical Review E*, vol. 70, no. 4, p. 041904, oct 2004.

[105] A. Franci, A. Chaillet, and W. Pasillas-Lépine, "Existence and robustness of phase-locking in coupled Kuramoto oscillators under mean-field feedback," *Automatica*, vol. 47, no. 6, pp. 1193–1202, jun 2011.

[106] O. V. Popovych, B. Lysyansky, and P. A. Tass, "Closed-loop deep brain stimulation by pulsatile delayed feedback with increased gap between pulse phases," *Scientific Reports*, vol. 7, no. 1, p. 1033, dec 2017.

[107] O. V. Popovych, B. Lysyansky, M. Rosenblum, A. Pikovsky, P. A. Tass, and L. Zrinzo, "Pulsatile desynchronizing delayed feedback for closed-loop deep brain stimulation," *PLOS ONE*, vol. 12, no. 3, p. e0173363, mar 2017.

[108] D. T. Brocker, B. D. Swan, R. Q. So, D. A. Turner, R. E. Gross, and W. M. Grill, "Optimized temporal pattern of brain stimulation designed by computational evolution," *Science Translational Medicine*, vol. 9, no. 371, 2017.

[109] D. T. Brocker, B. D. Swan, D. A. Turner, R. E. Gross, S. B. Tatter, M. M. Koop, H. Bronte-Stewart, and W. M. Grill, "Improved efficacy of temporally non-regular deep brain stimulation in Parkinson's disease." *Experimental neurology*, vol. 239, pp. 60–7, jan 2013.

[110] P. Gorzelic, S. J. Schiff, and A. Sinha, "Model-based rational feedback controller design for closed-loop deep brain stimulation of Parkinson's disease," *Journal of Neural Engineering*, vol. 10, no. 2, p. 026016, apr 2013.

[111] A. Pascual, J. Modolo, and A. Beuter, "Is a computational model useful to understand the effect of deep brain stimulation in Parkinson's disease?" *Journal of integrative neuroscience*, vol. 5, no. 4, pp. 541–59, dec 2006.

[112] A. Priori, "Technology for deep brain stimulation at a gallop," *Movement Disorders*, vol. 30, no. 9, pp. 1206–1212, aug 2015.

[113] J. Lee, H. G. Rhew, D. R. Kipke, and M. P. Flynn, "A 64 channel programmable closed-loop neurostimulator with 8 channel neural amplifier and logarithmic ADC," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1935–1945, 2010.

[114] H.-g. Rhew, J. Jeong, J. A. Fredenburg, S. Member, S. Dodani, P. G. Patil, M. P. Flynn, and S. Member, "A Fully Self-Contained Logarithmic Closed-Loop Deep Brain Stimulation SoC With Wireless Telemetry and Wireless Power Management," pp. 1–15, 2014.

[115] S. Stanslaski, P. Afshar, P. Cong, J. Giftakis, P. Stypulkowski, D. Carlson, D. Linde, D. Ullestad, A.-T. Avestruz, and T. Denison, "Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 4, pp. 410–21, jul 2012.

[116] M. T. Salam, J. Luis, P. Velazquez, R. Genov, and S. Member, "Seizure Suppression Efficacy of Closed-Loop Versus Open-Loop Deep Brain Stimulation in a Rodent Model of Epilepsy," vol. 24, no. 6, pp. 710–719, 2016.

[117] M. Rosa, S. Marceglia, D. Servello, G. Foffani, L. Rossi, M. Sassi, S. Mrakic-Sposta, R. Zangaglia, C. Pacchetti, M. Porta, and A. Priori, "Time dependent subthalamic local field potential changes after DBS surgery in Parkinson's disease," *Experimental Neurology*, vol. 222, no. 2, pp. 184–190, apr 2010.

[118] G. Deuschl, J. Herzog, G. Kleiner-Fisman, C. Kubu, A. M. Lozano, K. E. Lyons, M. C. Rodriguez-Oroz, F. Tamma, A. I. Tröster, J. L. Vitek, J. Volkmann, and V. Voon, "Deep brain stimulation: Postoperative issues," *Movement Disorders*, vol. 21, no. S14, pp. S219–S237, jun 2006.

[119] M. K. Hosain, A. Kouzani, and S. Tye, "Closed loop deep brain stimulation: an evolving technology," *Australasian Physical & Engineering Sciences in Medicine*, vol. 37, no. 4, pp. 619–634, 2014.

[120] S. Little, A. Pogosyan, A. A. Kuhn, and P. Brown, "β band stability over time correlates with Parkinsonian rigidity and bradykinesia." *Experimental neurology*, vol. 236, no. 2, pp. 383–8, aug 2012.

[121] M. Alegre and M. Valencia, "Oscillatory activity in the human basal ganglia: more than just beta, more than just Parkinson's disease." *Experimental neurology*, vol. 248, pp. 183–6, oct 2013.

[122] J.-S. Brittain and P. Brown, *NeuroImage*, vol. 85 Pt 2, pp. 637–47, jan 2014.

[123] P. Brown and D. Williams, "Basal ganglia local field potential activity: Character and functional significance in the human," *Clinical Neurophysiology*, vol. 116, no. 11, pp. 2510–2519, 2005.

[124] T. Heida, E. C. Wentink, and E. Marani, "Power spectral density analysis of physiological, rest and action tremor in Parkinson's disease patients treated with deep brain stimulation." *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 70, jan 2013.

[125] M. F. Contarino, L. J. Bour, R. Verhagen, M. A. J. Lourens, R. M. A. de Bie, P. van den Munckhof, and P. R. Schuurman, "Directional steering: A novel approach to deep brain stimulation." *Neurology*, vol. 83, no. 13, pp. 1163–9, sep 2014.

[126] A. Chaturvedi, T. J. Foutz, and C. C. McIntyre, "Current steering to activate targeted neural pathways during deep brain stimulation of the subthalamic region," *Brain Stimulation*, vol. 5, no. 3, pp. 369–377, jul 2012.

[127] V. Valente, A. Demosthenous, and R. Bayford, "A tripolar current-steering stimulator ASIC for field shaping in deep brain stimulation," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 3, pp. 197–207, 2012.

[128] ——, "Output stage of a current-steering multipolar and multisite deep brain stimulator," in *2013 IEEE Biomedical Circuits and Systems Conference, BioCAS 2013*. IEEE, 2013, pp. 85–88.

[129] A. Mohammed, V. Valente, R. Bayford, and A. Demosthenous, "Output stage of a dynamic current steering deep brain stimulator," in *2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*. IEEE, dec 2015, pp. 81–84.

[130] M. Keane, S. Deyo, A. Abosch, J. A. Bajwa, and M. D. Johnson, "Improved spatial targeting with directionally segmented deep brain stimulation leads for treating essential tremor." *Journal of neural engineering*, vol. 9, no. 4, p. 046005, aug 2012.

[131] C. Pollo, A. Kaelin-Lang, M. F. Oertel, L. Stieglitz, E. Taub, P. Fuhr, A. M. Lozano, A. Raabe, and M. Schupbach, "Directional deep brain stimulation: an intraoperative double-blind pilot study," *Brain*, vol. 137, no. 7, pp. 2015–2026, jul 2014.

[132] A. T. Connolly, R. J. Vetter, J. F. Hetke, B. A. Teplitzky, D. R. Kipke, D. S. Pellinen, D. J. Anderson, K. B. Baker, J. L. Vitek, and M. D. Johnson, "A Novel Lead Design for Modulation and Sensing of Deep Brain Structures." *IEEE transactions on bio-medical engineering*, vol. 63, no. 1, pp. 148–157, jan 2016.

[133] R. E. Gross, P. Krack, M. C. Rodriguez-Oroz, A. R. Rezai, and A.-L. Benabid, "Electrophysiological mapping for the implantation of deep brain stimulators for Parkinson's disease and tremor," *Movement Disorders*, vol. 21, no. S14, pp. S259–S283, jun 2006.

[134] M. L. Janssen, D. G. Zwartjes, Y. Temel, V. van Kranen-Mastenbroek, A. Duits, L. J. Bour, P. H. Veltink, T. Heida, and V. Visser-Vandewalle, "Subthalamic neuronal responses to cortical stimulation," *Movement Disorders*, vol. 27, no. 3, pp. 435–438, mar 2012.

[135] R. J. Coffey, "Deep Brain Stimulation Devices: A Brief Technical History and Review," *Artificial Organs*, vol. 33, no. 3, pp. 208–220, mar 2009.

[136] J. Gardner, "A history of deep brain stimulation: Technological innovation and the role of clinical assessment tools," *Social Studies of Science*, vol. 43, no. 5, pp. 707–728, oct 2013.

[137] I. Ughratdar, M. Samuel, and K. Ashkan, "Technological Advances in Deep Brain Stimulation," *Journal of Parkinson's Disease*, vol. 5, no. 3, pp. 483–496, sep 2015.

[138] J. Dams, U. Siebert, B. Bornschein, J. Volkmann, G. Deuschl, W. H. Oertel, R. Dodel, and J.-P. Reese, "Cost-effectiveness of deep brain stimulation in patients with Parkinson's disease," *Movement Disorders*, vol. 28, no. 6, pp. 763–771, jun 2013.

[139] E. R. Dorsey, R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf, and C. M. Tanner, "Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030," *Neurology*, no. 5, pp. 384–386, jan.

[140] Transparency Market Research, "Global Deep Brain Stimulation Devices Market for Parkinsons Disease to be Worth US$9.4 bn by 2019 Driven by Rising Aging Population," 2016. [Online]. Available: http://www.transparencymarketresearch.com/pressrelease/deep-brain-stimulator-market.htm

[141] C. K. E. Moll and A. K. Engel, "Phase matters: cancelling pathological tremor by adaptive deep brain stimulation," *Brain*, vol. 140, no. 1, pp. 5–8, jan 2017.

[142] R. E. Elliott, A. Morsi, O. Tanweer, B. Grobelny, E. Geller, C. Carlson, O. Devinsky, and W. K. Doyle, "Efficacy of vagus nerve stimulation over time: Review of 65 consecutive patients with treatment-resistant epilepsy treated with VNS greater than 10years," *Epilepsy & Behavior*, vol. 20, no. 3, pp. 478–483, mar 2011.

[143] S. Mason and G. Birch, "A general framework for brain-computer interface design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 1, pp. 70–85, mar 2003.

[144] M. A. Lebedev and M. A. L. Nicolelis, "Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation," *Physiological Reviews*, vol. 97, no. 2, 2017.

[145] M. A. Nicolelis, "Actions from thoughts." *Nature*, vol. 409, no. 6818, pp. 403–407, jan 2001.

[146] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Braincomputer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, pp. 513–525, aug 2016.

[147] D. Wu, B. J. Lance, and V. J. Lawhern, "Guest Editorial for the Special Section on Brain Computer Interface (BCI)," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 1–2, feb 2017.

[148] S. Ha, A. Akinin, J. Park, C. Kim, H. Wang, C. Maier, P. P. Mercier, and G. Cauwenberghs, "Silicon-Integrated High-Density Electrocortical Interfaces," *Proceedings of the IEEE*, vol. 105, no. 1, pp. 11–33, jan 2017.

[149] J. Muraskin, J. Sherwin, G. Lieberman, J. O. Garcia, T. Verstynen, J. M. Vettel, and P. Sajda, "Fusing Multiple Neuroimaging Modalities to Assess Group Differences in PerceptionAction Coupling," *Proceedings of the IEEE*, vol. 105, no. 1, pp. 83–100, jan 2017.

[150] M. Mills, "Hearing Aids and the History of Electronics Miniaturization," *IEEE Annals of the History of Computing*, vol. 33, no. 2, pp. 24–45, feb 2011.

[151] L. Greenemeier, "FDA Approves First Retinal Implant," feb 2013. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature.2013.12439

[152] J. A. Costa e Silva and R. E. Steffen, "The future of psychiatry: brain devices," *Metabolism*, vol. 69, pp. S8–S12, April 2017.

[153] Y. H.-L. Luo and L. da Cruz, "The Argus II Retinal Prosthesis System," *Progress in Retinal and Eye Research*, vol. 50, pp. 89–107, jan 2016.

[154] A. L. HODGKIN and A. F. HUXLEY, "Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo." *The Journal of physiology*, vol. 116, no. 4, pp. 449–72, apr 1952.

[155] D. H. Hubel and T. N. Wiesel, "Ferrier Lecture: Functional Architecture of Macaque Monkey Visual Cortex," *Proceedings of the Royal Society B: Biological Sciences*, vol. 198, no. 1130, pp. 1–59, may 1977.

[156] R. J. Staba, C. L. Wilson, A. Bragin, I. Fried, and J. Engel, "Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex." *Journal of neurophysiology*, vol. 88, no. 4, pp. 1743–52, oct 2002.

[157] G. G. Yener and E. Baar, "Brain oscillations as biomarkers in neuropsychiatric disorders: following an interactive panel discussion and synopsis." *Supplements to Clinical neurophysiology*, vol. 62, pp. 343–63, jan 2013.

[158] R. Brette, Z. Piwkowska, C. Monier, M. Rudolph-Lilith, J. Fournier, M. Levy, Y. Frégnac, T. Bal, and A. Destexhe, "High-resolution intracellular recordings using a real-time computational model of the electrode." *Neuron*, vol. 59, no. 3, pp. 379–91, aug 2008.

[159] R. Brette and A. Destexhe, Eds., *Handbook of Neural Activity Measurement*. Cambridge, UK: Cambridge University Press, 2012.

[160] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Computation in Neural Systems*, vol. 9, no. 4, pp. 53–78, jul 1998.

[161] C. Gold, D. A. Henze, C. Koch, and G. Buzsáki, "On the origin of the extracellular action potential waveform: A modeling study." *Journal of neurophysiology*, vol. 95, no. 5, pp. 3113–28, may 2006.

[162] G. Buzsaki, "Large-scale recording of neuronal ensembles." *Nature neuroscience*, vol. 7, no. 5, pp. 446–51, may 2004.

[163] A. Priori, G. Foffani, L. Rossi, and S. Marceglia, "Adaptive deep brain stimulation (aDBS) controlled by local field potential oscillations," *Experimental Neurology*, vol. 245, pp. 77–86, 2012.

[164] E. Urrestarazu, J. Iriarte, M. Alegre, P. Clavero, M. C. Rodríguez-Oroz, J. Guridi, J. A. Obeso, and J. Artieda, "Beta activity in the subthalamic nucleus during sleep in patients with Parkinson's disease." *Movement disorders : official journal of the Movement Disorder Society*, vol. 24, no. 2, pp. 254–60, jan 2009.

[165] A. A. Kuhn, A. Tsui, T. Aziz, N. Ray, C. Brücke, A. Kupsch, G.-H. Schneider, and P. Brown, "Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity." *Experimental neurology*, vol. 215, no. 2, pp. 380–7, feb 2009.

[166] N. Rowland, J. Breshears, and E. Chang, "Neurosurgery and the dawning age of Brain-Machine Interfaces," *Surgical Neurology International*, vol. 4, no. 2, p. 11, 2013.

[167] A. B. Schwartz, X. T. Cui, D. Weber, and D. W. Moran, "Brain-Controlled Interfaces: Movement Restoration with Neural Prosthetics," *Neuron*, vol. 52, no. 1, pp. 205–220, oct 2006.

[168] G. Schalk and E. C. Leuthardt, "Brain-Computer Interfaces Using Electrocorticographic Signals," *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 140–154, 2011.

[169] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors (Basel, Switzerland)*, vol. 12, no. 2, pp. 1211–1279, jan 2012.

[170] S. K. B. Bick, B. S. Folley, J. S. Mayer, S. Park, P. D. Charles, C. R. Camalier, S. Pallavaram, P. E. Konrad, and J. S. Neimat, "Subthalamic Nucleus Deep Brain Stimulation Alters Prefrontal Correlates of Emotion Induction," *Neuromodulation: Technology at the Neural Interface*, vol. 20, no. 3, pp. 233–237, apr 2017.

[171] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nature Methods*, vol. 2, no. 12, pp. 932–940, dec 2005.

[172] H. Takehara, Y. Ohta, M. Motoyama, M. Haruta, M. Nagasaki, H. Takehara, T. Noda, K. Sasagawa, T. Tokuda, and J. Ohta, "Intravital fluorescence imaging of mouse brain using implantable semiconductor devices and epi-illumination of biological tissue." *Biomedical optics express*, vol. 6, no. 5, pp. 1553–64, may 2015.

[173] S. R. Schultz, C. S. Copeland, A. J. Foust, P. Quicke, and R. Schuck, "Advances in Two-Photon Scanning and Scanless Microscopy Technologies for Functional Neural Circuit Imaging," *Proceedings of the IEEE*, vol. 105, no. 1, pp. 139–157, jan 2017.

[174] M. J. Berridge, P. Lipp, and M. D. Bootman, "The versatility and universality of calcium signalling." *Nature reviews. Molecular cell biology*, vol. 1, no. 1, pp. 11–21, oct 2000.

[175] M. T. Alt, E. Fiedler, L. Rudmann, J. S. Ordonez, P. Ruther, and T. Stieglitz, "Let There Be LightOptoprobes for Neural Implants," *Proceedings of the IEEE*, vol. 105, no. 1, pp. 101–138, jan 2017.

[176] S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain-computer interface using a simplified functional near-infrared spectroscopy system." *Journal of neural engineering*, vol. 4, no. 3, pp. 219–226, sep 2007.

[177] K. Sakatani, Y. Katayama, T. Yamamoto, and S. Suzuki, "Changes in cerebral blood oxygenation of the frontal lobe induced by direct electrical stimulation of thalamus and globus pallidus: a near infrared spectroscopy study." *Journal of neurology, neurosurgery, and psychiatry*, vol. 67, no. 6, pp. 769–73, dec 1999.

[178] T. Morishita, M.-A. Higuchi, K. Saita, Y. Tsuboi, H. Abe, and T. Inoue, "Changes in Motor-Related Cortical Activity Following Deep Brain Stimulation for Parkinson's Disease Detected by Functional Near Infrared Spectroscopy: A Pilot Study." *Frontiers in human neuroscience*, vol. 10, p. 629, 2016.

[179] E. Fouragnan, C. Retzler, K. Mullinger, and M. G. Philiastides, "Two spatiotemporally distinct value systems shape reward-based learning in the human brain," *Nature Communications*, vol. 6, p. 8107, sep 2015.

[180] P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli, "Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond." *Radiographics : a review publication of the Radiological Society of North America, Inc*, vol. 26 Suppl 1, pp. S205–23, oct 2006.

[181] R. L. Muetzel, P. F. Collins, B. A. Mueller, A. M. Schissel, K. O. Lim, and M. Luciana, "The development of corpus callosum microstructure and associations with bimanual task performance in healthy adolescents," *NeuroImage*, vol. 39, no. 4, pp. 1918–1925, feb 2008.

[182] K. H. Lee, K. Kristic, R. van Hoff, F. L. Hitti, C. Blaha, B. Harris, D. W. Roberts, and J. C. Leiter, "High-frequency stimulation of the subthalamic nucleus increases glutamate in the subthalamic nucleus of rats as demonstrated by in vivo enzyme-linked glutamate sensor." *Brain research*, vol. 1162, pp. 121–129, aug 2007.

[183] D. L. Robinson, B. J. Venton, M. L. A. V. Heien, and R. M. Wightman, "Detecting subsecond dopamine release with fast-scan cyclic voltammetry in vivo." *Clinical chemistry*, vol. 49, no. 10, pp. 1763–1773, oct 2003.

[184] T. Morishita and T. Inoue, "Need for multiple biomarkers to adjust parameters of closed-loop deep brain stimulation for Parkinson's disease." *Neural regeneration research*, vol. 12, no. 5, pp. 747–748, may 2017.

[185] S. A. Kim and S. B. Jun, "In-vivo Optical Measurement of Neural Activity in the Brain." *Experimental neurobiology*, vol. 22, no. 3, pp. 158–66, sep 2013.

[186] J. Jorge, W. van der Zwaag, and P. Figueiredo, "EEGfMRI integration for the study of human brain function," *NeuroImage*, vol. 102, pp. 24–34, nov 2014.

[187] W. H. P. Secretary, "Fact sheet: BRAIN Initiative," 2013. [Online]. Available: https://obamawhitehouse.archives.gov/the-press-office/2013/04/02/fact-sheet-brain-initiative

[188] S. Marceglia, E. Rossi, M. Rosa, F. Cogiamanian, L. Rossi, L. Bertolasi, A. Vogrig, F. Pinciroli, S. Barbieri, and A. Priori, "Web-based telemonitoring and delivery of caregiver support for patients with Parkinson disease after deep brain stimulation: protocol." *JMIR research protocols*, vol. 4, no. 1, p. e30, mar 2015.

[189] D. D. French, R. R. Campbell, S. Sabharwal, A. L. Nelson, P. A. Palacios, and D. Gavin-Dreschnack, "Health care costs for patients with chronic spinal cord injury in the Veterans Health Administration." *The journal of spinal cord medicine*, vol. 30, no. 5, pp. 477–81, 2007.

[190] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview." *Bioinformatics (Oxford, England)*, vol. 16, no. 5, pp. 412–424, may 2000.

[191] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, may 2011.

[192] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, jun 2006.

[193] M. Fatourechi, R. K. Ward, and G. E. Birch, "A self-paced braincomputer interface system with a low false positive rate," *Journal of Neural Engineering*, vol. 5, no. 1, pp. 9–23, mar 2008.

[194] A. G. Rouse, S. R. Stanslaski, P. Cong, R. M. Jensen, P. Afshar, D. Ullestad, R. Gupta, G. F. Molnar, D. W. Moran, and T. J. Denison, "A chronic generalized bi-directional brainmachine interface," *Journal of Neural Engineering*, vol. 8, no. 3, p. 036018, jun 2011.

[195] S.-S. Yoo, H. Kim, E. Filandrianos, S. J. Taghados, and S. Park, "Non-Invasive Brain-to-Brain Interface (BBI): Establishing Functional Links between Two Brains," *PLoS ONE*, vol. 8, no. 4, p. e60410, apr 2013.

[196] T. Fukami, T. Shimada, E. Forney, and C. W. Anderson, "EEG character identification using stimulus sequences designed to maximize mimimal hamming distance," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, aug 2012, pp. 1782–1785.

[197] J. Goni, M. P. van den Heuvel, A. Avena-Koenigsberger, N. Velez de Mendizabal, R. F. Betzel, A. Griffa, P. Hagmann, B. Corominas-Murtra, J.-P. Thiran, and O. Sporns, "Resting-brain functional connectivity predicted by analytic measures of network communication." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 2, pp. 833–8, jan 2014.

[198] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual Image Reconstruction from Human Brain

Activity using a Combination of Multiscale Local Image Decoders," *Neuron*, vol. 60, no. 5, pp. 915–929, dec 2008.

[199] A. Akce, J. J. S. Norton, and T. Bretl, "An SSVEP-Based Brain-Computer Interface for Text Spelling With Adaptive Queries That Maximize Information Gain Rates," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 857–866, sep 2015.

[200] G. Schalk, P. Brunner, L. Gerhardt, H. Bischof, and J. Wolpaw, "Braincomputer interfaces (BCIs): Detection instead of classification," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 51–62, jan 2008.

[201] E. Yargholi and G.-A. Hossein-Zadeh, "Brain Decoding-Classification of Hand Written Digits from fMRI Data Employing Bayesian Networks." *Frontiers in human neuroscience*, vol. 10, p. 351, 2016.

[202] H. Skutkova, M. Vitek, K. Sedlar, and I. Provaznik, "Progressive alignment of genomic signals by multiple dynamic time warping," *Journal of Theoretical Biology*, vol. 385, pp. 20–30, nov 2015.

[203] K. H. Lee, S.-Y. Kung, and N. Verma, "Low-energy Formulations of Support Vector Machine Kernel Functions for Biomedical Sensor Applications," *Journal of Signal Processing Systems*, vol. 69, no. 3, pp. 339–349, dec 2012.

[204] H. Kassiri, S. Tonekaboni, M. T. Salam, N. Soltani, K. Abdelhalim, J. L. P. Velazquez, and R. Genov, "Closed-Loop Neurostimulators: A Survey and A Seizure-Predicting Design Example for Intractable Epilepsy Treatment," *IEEE Transactions on Biomedical Circuits and Systems*, pp. 1–15, 2017.

[205] H. Kassiri, A. Bagheri, N. Soltani, K. Abdelhalim, H. M. Jafari, M. T. Salam, J. L. Perez Velazquez, and R. Genov, "Battery-less Tri-band-Radio Neuro-monitor and Responsive Neurostimulator for Diagnostics and Treatment of Neurological Disorders," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 5, pp. 1274–1289, may 2016.

[206] W. Biederman, D. J. Yeager, N. Narevsky, J. Leverett, R. Neely, J. M. Carmena, E. Alon, and J. M. Rabaey, "A 4.78 mm 2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels With Digital Compression and Simultaneous Dual Stimulation," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 1038–1047, apr 2015.

[207] R. Shulyzki, K. Abdelhalim, A. Bagheri, M. T. Salam, C. M. Florez, J. L. P. Velazquez, P. L. Carlen, and R. Genov, "320-Channel Active Probe for High-Resolution Neuromonitoring and Responsive Neurostimulation," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 1, pp. 34–49, feb 2015.

[208] W.-M. Chen, H. Chiueh, T.-J. Chen, C.-L. Ho, C. Jeng, M.-D. Ker, C.-Y. Lin, Y.-C. Huang, C.-W. Chou, T.-Y. Fan, M.-S. Cheng, Y.-L. Hsin, S.-F. Liang, Y.-L. Wang, F.-Z. Shaw, Y.-H. Huang, C.-H. Yang, and C.-Y. Wu, "A Fully Integrated 8-Channel Closed-Loop Neural-Prosthetic CMOS SoC for Real-Time Epileptic Seizure Control," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, jan 2014.

[209] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential braincomputer interface (BCI): The effects

of matrix size and inter stimulus interval on performance," *Biological Psychology*, vol. 73, no. 3, pp. 242–252, 2006.

[210] X. Liu, M. Zhang, A. G. Richardson, T. H. Lucas, and J. Van der Spiegel, "Design of a Closed-Loop, Bidirectional Brain Machine Interface System With Energy Efficient Neural Feature Extraction and PID Control," *IEEE Transactions on Biomedical Circuits and Systems*, pp. 1–14, 2016.

[211] D. Markovic and R. Brodersen, *DSP Architecture Design Essentials.* New York, NY: Springer, 2012.

[212] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits - A Design Perspective*, 2nd ed. Pearson, 2002.

[213] I. Pisotta, D. Perruchoud, and S. Ionta, "Hand-in-hand advances in biomedical engineering and sensorimotor restoration," *Journal of Neuroscience Methods*, vol. 246, pp. 22–29, may 2015.

[214] M. Kilpatrick, E. Church, S. Danish, M. Stiefel, J. Jaggi, C. Halpern, M. Kerr, E. Maloney, M. Robinson, I. Lucki, E. Krizman-Grenda, and G. Baltuch, "Intracerebral microdialysis during deep brain stimulation surgery," *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 106–111, jun 2010.

[215] T. Endo, R. Okuno, M. Yokoe, K. Akazawa, and S. Sakoda, "A novel method for systematic analysis of rigidity in Parkinson's disease." *Movement disorders : official journal of the Movement Disorder Society*, vol. 24, no. 15, pp. 2218–24, nov 2009.

[216] M. Rivlin-Etzion, O. Marmor, G. Heimer, A. Raz, A. Nini, and H. Bergman, "Basal ganglia oscillations and pathophysiology of movement disorders." *Current opinion in neurobiology*, vol. 16, no. 6, pp. 629–37, dec 2006.

[217] J. F. Baizabal-Carvallo and M. Alonso-Juarez, "Low-frequency deep brain stimulation for movement disorders," *Parkinsonism & Related Disorders*, 2016.

[218] S. Marceglia, M. Fumagalli, and A. Priori, "What neurophysiological recordings tell us about cognitive and behavioral functions of the human subthalamic nucleus." *Expert review of neurotherapeutics*, vol. 11, no. 1, pp. 139–49, jan 2011.

[219] R. Katzenschlager and A. J. Lees, "Treatment of Parkinson's disease: levodopa as the first choice." *Journal of neurology*, vol. 249 Suppl, pp. II19–24, sep 2002.

[220] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease." *Progress in neurobiology*, vol. 81, no. 1, pp. 29–44, jan 2007.

[221] G. Birot, L. Albera, F. Wendling, and I. Merlet, "Localization of extended brain sources from EEG/MEG: The ExSo-MUSIC approach," *NeuroImage*, vol. 56, no. 1, pp. 102–113, may 2011.

[222] A. Hyvarinen, "Testing the ICA mixing matrix based on inter-subject or inter-session consistency," *NeuroImage*, vol. 58, no. 1, pp. 122–136, sep 2011.

[223] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Boca Raton, Florida: Chapman & Hall/CRC, 2004.

[224] J. Muthuswamy, "Biomedical Signal Analysis," in *Standard Handbook of Biomedical Engineering and Design*, M. Kutz, Ed. The McGraw-Hill Companies, 2003, ch. 18, pp. 1 – 30.

[225] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Zhilin Zhang, and K. Kreutz-Delgado, "Evolving Signal Processing for BrainComputer Interfaces," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, may 2012.

[226] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. New Jersey: John Wiley & Sons, Inc., 2008.

[227] J. Semmlow, *Signals and Systems for Bioengineers*. Elsevier, 2012.

[228] G. Reinoso, J. C. Allen, W.-L. Au, S.-H. Seah, K.-Y. Tay, and L. C. S. Tan, "Clinical evolution of Parkinson's disease and prognostic factors affecting motor progression: 9-year follow-up study." *European journal of neurology*, vol. 22, no. 3, pp. 457–63, mar 2015.

[229] C. Ding, G. Ganesvaran, B. G. Clissold, C. D. Mccoll, K. A. Reardon, and M. Schiff, "Study of Levodopa Response in Parkinson s Disease : Observations on Rates of Motor Progression," vol. 00, no. 00, pp. 1–4, 2016.

[230] C. Walck, "Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists," 1996. [Online]. Available: http://www.stat.rice.edu/{~}dobelman/textfiles/DistributionsHandbook.pdf

[231] S. Gibson, J. W. Judy, and D. Marković, "Technology-aware algorithm design for neural spike detection, feature extraction, and dimensionality reduction." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 18, no. 5, pp. 469–478, 2010.

[232] P. Prandoni and M. Vetterli., *Signal processing for Communications*, first edit ed. Boca Raton, FL: CRC Press, 2008.

[233] D. Sundararajan, *Discrete Wavelet Transform: A Signal Processing Approach*. New York, United States: John Wiley and Sons Ltd, 2015.

[234] C. J. C. Burges, "Dimension Reduction: A Guided Tour," *Foundations and Trends in Machine Learning*, vol. 2, no. 4, pp. 275–365, 2009.

[235] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd ed. San Diego: Academic Press, 2006.

[236] M. Barni, F. Buti, F. Bartolini, and V. Cappellini, "A quasi-Euclidean norm to speed up vector median filtering." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 10, pp. 1704–9, jan 2000.

[237] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, UK: Cambridge University Press, 2012.

[238] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000.

[239] M. Zamani and A. Demosthenous, "Feature Extraction Using Extrema Sampling of Discrete Derivatives for Spike Sorting in Implantable Upper-Limb Neural Prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 716–726, jul 2014.

[240] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, mar 2003.

[241] S. J. Schiff, "Towards model-based control of Parkinson's disease." *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 368, no. 1918, pp. 2269–308, may 2010.

[242] M. H. Myers, M. Threatt, K. M. Solies, B. M. McFerrin, L. B. Hopf, J. D. Birdwell, and K. A. Sillay, "Ambulatory Seizure Monitoring: From Concept to Prototype Device." *Annals of neurosciences*, vol. 23, no. 2, pp. 100–11, jul 2016.

[243] E. Rossi, M. Rosa, L. Rossi, A. Priori, and S. Marceglia, "WebBioBank: A new platform for integrating clinical forms and shared neurosignal analyses to support multi-centre studies in Parkinsons Disease," *Journal of Biomedical Informatics*, vol. 52, pp. 92–104, 2014.

[244] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. P. Chandrakasan, "A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 804–816, apr 2010.

[245] M. Ciletti, *Advanced Digital Design with Verilog HDL*, 2nd ed. New Jersey: Pearson Higher Education, 2010.

[246] T. Cormen, *Introduction to Algorithms*, 3rd ed. Cambridge, Massachusetts: MIT Press, 2009.

[247] A. Page, C. Sagedy, E. Smith, N. Attaran, T. Oates, and T. Mohsenin, "A Flexible Multichannel EEG Feature Extractor and Classifier for Seizure Detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 109–113, feb 2015.

[248] M. Hasanlou and F. Samadzadegan, "Comparative Study of Intrinsic Dimensionality Estimation and Dimension Reduction Techniques on Hyperspectral Images Using K-NN Classifier," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1046–1050, nov 2012.

[249] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[250] K. Zhou, H. Qiming, and B. Guo, "Real-time kd-tree construction on graphics hardware," 2013.

[251] Xilinx, "7 Series FPGA Data Sheet: Overview," p. 18, 2017. [Online]. Available: https://www.xilinx.com/support/documentation/data{_}sheets/ds180{_}7Series{_}Overview.pdf

[252] Digilent, "Nexys4 DDR FPGA Board Refrence Manual," p. 29, 2016. [Online]. Available: https://reference.digilentinc.com/{_}media/reference/programmable-logic/nexys-4-ddr/nexys4ddr{_}rm.pdf

[253] Nangate Inc, "NanGate FreePDK45 Open Cell Library," 2017. [Online]. Available: http://www.nangate.com/?page{_}id=2325

[254] S. Iranmanesh and E. Rodriguez-Villegas, "A 950 nW Analog-Based Data Reduction Chip for Wearable EEG Systems in Epilepsy," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2362–2373, sep 2017.

[255] M. Shoaran, C. Pollo, K. Schindler, and A. Schmid, "A Fully Integrated IC With 0.85 uW/Channel Consumption for Epileptic iEEG Detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 114–118, feb 2015.

[256] M. Shoaib, K. H. Lee, N. K. Jha, and N. Verma, "A 0.6107 uW Energy-Scalable Processor for Directly Analyzing Compressively-Sensed EEG," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 4, pp. 1105–1118, apr 2014.

[257] M. A. Bin Altaf and J. Yoo, "A 1.83 J/Classification, 8-Channel, Patient-Specific Epileptic Seizure Classification SoC Using a Non-Linear Support Vector Machine," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 1, pp. 49–60, feb 2016.

[258] M. A. Bin Altaf, C. Zhang, and J. Yoo, "A 16-Channel Patient-Specific Seizure Onset and Termination Detection SoC With Impedance-Adaptive Transcranial Electrical Stimulator," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, nov 2015.

[259] T.-J. Chen, S.-C. Lee, C.-H. Yang, C.-F. Chiu, and H. Chiueh, "A 28.6 uW mixed-signal processor for epileptic seizure detection," in *2013 Symposium on VLSI Circuits (VLSIC)*. Kyoto, Japan: IEEE, 2013, pp. C52–C53.

[260] M. Parastarfeizabadi, A. Z. Kouzani, M. Moffitt, K. Otto, D. Kipke, and C. McIntyre, "Advances in closed-loop deep brain stimulation devices," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 79, dec 2017.

[261] A. E. W. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. Clifton, and G. D. Clifford, "Machine Learning and Decision Support in Critical Care," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 444–466, feb 2016.

[262] M. Sung, C. Marci, and A. Pentland, "Wearable feedback systems for rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, p. 17, jun 2005.

[263] P. J. Soh, G. A. Vandenbosch, M. Mercuri, and D. M.-P. Schreurs, "Wearable Wireless Health Monitoring: Current Developments, Challenges, and Future Trends," *IEEE Microwave Magazine*, vol. 16, no. 4, pp. 55–70, may 2015.

[264] A. Zhang, L. Wang, X. Ye, and X. Lin, "Light-Weight and Robust Security-Aware D2D-Assist Data Transmission Protocol for Mobile-Health Systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 662–675, mar 2017.

[265] M. Picillo, A. M. Lozano, N. Kou, R. Puppi Munhoz, and A. Fasano, "Programming Deep Brain Stimulation for Parkinson's Disease: The Toronto Western Hospital Algorithms," *Brain Stimulation*, vol. 9, no. 3, pp. 425–437, 2016.

[266] S. Zavitsanou, A. Chakrabarty, E. Dassau, and F. Doyle, "Embedded Control in Wearable Medical Devices: Application to the Artificial Pancreas," *Processes*, vol. 4, no. 4, p. 35, sep 2016.

[267] K. Zarkogianni, A. Vazeou, S. G. Mougiakakou, A. Prountzou, and K. S. Nikita, "An Insulin Infusion Advisory System Based on Autotuning Nonlinear Model-Predictive Control," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2467–2477, sep 2011.

[268] K. Soltesz, J.-O. Hahn, T. Hägglund, G. A. Dumont, and J. M. Ansermino, "Individualized closed-loop control of propofol anesthesia: A preliminary study," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 500–508, 2013.

[269] R. Barro, Senen, Marin, Ed., *Fuzzy Logic in Medicine*. Heidelberg: Springer Science & Business Media, 2002.

[270] P. Sajda, "Machine Learning for Detection and Diagnosis of Disease," *Annual Review of Biomedical Engineering*, vol. 8, no. 1, pp. 537–565, aug 2006.

[271] A. Csavoy, G. Molnar, and T. Denison, "Creating support circuits for the nervous system: Considerations for brain-machine interfacing," in *IEEE Symposium on VLSI Circuits*. IEEE, 2009, pp. 4–7.

[272] A. B. Rossow, E. O. T. Salles, and K. F. Coco, "Automatic sleep staging using a single-channel EEG modeling by Kalman Filter and HMM," in *ISSNIP Biosignals and Biorobotics Conference 2011*. IEEE, jan 2011, pp. 1–6.

[273] Z. H. T. Kiss, D. M. Mooney, L. Renaud, and B. Hu, "Neuronal response to local electrical stimulation in rat thalamus: physiological implications for mechanisms of deep brain stimulation." *Neuroscience*, vol. 113, no. 1, pp. 137–43, 2002.

[274] M. J. Birdno and W. M. Grill, "Mechanisms of deep brain stimulation in movement disorders as revealed by changes in stimulus frequency." *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, vol. 5, no. 1, pp. 14–25, jan 2008.

[275] A. D. Dorval, A. M. Kuncel, M. J. Birdno, D. A. Turner, and W. M. Grill, "Deep brain stimulation alleviates parkinsonian bradykinesia by regularizing pallidal activity." *Journal of neurophysiology*, vol. 104, no. 2, pp. 911–21, aug 2010.

[276] A. M. Kuncel and W. M. Grill, "Selection of stimulus parameters for deep brain stimulation," *Clinical Neurophysiology*, vol. 115, no. 11, pp. 2431–2441, 2004.

[277] C. Moreau, L. Defebvre, A. Destée, S. Bleuse, F. Clement, J. L. Blatt, P. Krystkowiak, and D. Devos, "STN-DBS frequency effects on freezing of gait in advanced Parkinson disease." *Neurology*, vol. 71, no. 2, pp. 80–4, jul 2008.

[278] A. Belasen, K. Rizvi, L. E. Gee, P. Yeung, J. Prusik, A. Ramirez-Zamora, E. Hanspal, P. Paiva, J. Durphy, C. E. Argoff, and J. G. Pilitsis, "Effect of low-frequency deep brain stimulation on sensory thresholds in Parkinson's disease," *Journal of Neurosurgery*, pp. 1–7, apr 2016.

[279] T. Xie, U. J. Kang, and P. Warnke, "Effect of stimulation frequency on immediate freezing of gait in newly activated STN DBS in Parkinson's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 83, no. 10, pp. 1015–1017, oct 2012.

[280] E. Moro, R. J. A. Esselink, J. Xie, M. Hommel, A. L. Benabid, and P. Pollak, "The impact on Parkinson's disease of electrical parameter settings in STN stimulation." *Neurology*, vol. 59, no. 5, pp. 706–13, sep 2002.

[281] A. Eusebio, W. Thevathasan, L. Doyle Gaynor, A. Pogosyan, E. Bye, T. Foltynie, L. Zrinzo, K. Ashkan, T. Aziz, and P. Brown, "Deep brain stimulation can suppress pathological synchronisation in parkinsonian patients," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 82, no. 5, pp. 569–573, may 2011.

[282] D. Whitmer, C. de Solages, B. Hill, H. Yu, J. M. Henderson, and H. Bronte-Stewart, "High frequency deep brain stimulation attenuates subthalamic and cortical rhythms in Parkinson's disease." *Frontiers in human neuroscience*, vol. 6, p. 155, 2012.

[283] K. B. Baker, J. Zhang, and J. L. Vitek, "Pallidal stimulation: effect of pattern and rate on bradykinesia in the non-human primate model of Parkinson's disease." *Experimental neurology*, vol. 231, no. 2, pp. 309–13, oct 2011.

[284] S. Vercruysse, W. Vandenberghe, L. Münks, B. Nuttin, H. Devos, and A. Nieuwboer, "Effects of deep brain stimulation of the subthalamic nucleus on freezing of gait in Parkinson's disease: a prospective controlled study." *Journal of neurology, neurosurgery, and psychiatry*, vol. 85, no. 8, pp. 871–7, aug 2014.

[285] G. Feng, "A Survey on Analysis and Design of Model-Based Fuzzy Control Systems," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 5, pp. 676–697, oct 2006.

[286] K. Passino and S. Yurkovich, *Fuzzy Control*, 1st ed. Boston, MA, USA: Addison Wesley Publishing Company, 1998.

[287] J. M. Bronstein, M. Tagliati, R. L. Alterman, A. M. Lozano, J. Volkmann, A. Stefani, F. B. Horak, M. S. Okun, K. D. Foote, P. Krack, R. Pahwa, J. M. Henderson, M. I. Hariz, R. A. Bakay, A. Rezai, W. J. Marks, E. Moro, J. L. Vitek, F. M. Weaver, R. E. Gross, and M. R. DeLong, "Deep Brain Stimulation for Parkinson Disease," *Archives of Neurology*, vol. 68, no. 2, p. 165, feb 2011.

[288] L. M. Zitella, K. Mohsenian, M. Pahwa, C. Gloeckner, and M. D. Johnson, "Computational modeling of pedunculopontine nucleus deep brain stimulation." *Journal of neural engineering*, vol. 10, no. 4, p. 045005, aug 2013.

[289] K. Sooksood, E. Noorsal, U. Bihr, and M. Ortmanns, "Recent advances in power efficient output stage for high density implantable stimulators," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 855–858, 2012.

[290] M. Ghovanloo and K. Najafi, "A compact large voltage-compliance high output-impedance programmable current source for implantable microstimulators," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 97–105, 2005.

[291] P. Hasgall, F. Di Gennaro, C. Baumgartner, E. Neufeld, M. Gosselin, D. Payne, A. Klingenbock, and N. Kuster, "ITIS Database for thermal and electromagnetic parameters of biological tissues," 2015.

[292] Boston Scientific, "Vercise Deep Brain Stimulator System," 2014. [Online]. Available: http://www.bostonscientific.com/en-EU/products/deep-brain-stimulation-systems/vercise-deep-brain-stimulation-system.html

Appendix A

Meaning of Mathematical Symbols and Notations

The symbols and notations have been defined in the main text of the thesis. However, they are defined again in a tabular form in this appendix for more clarity – related equations are grouped together. The symbols and notations are presented in the order they appear in each Chapter.

A.1 Symbols and Notations Used in Chapter 2

A.1.1 Equations 2.1 & 2.2

$E(t)$	Proportion of excitatory cells active per unit time.
$I(t)$	Proportion of inhibitory cells active per unit time.
$h_E(t)$	External input to the excitatory population.
$h_I(t)$	External input to the inhibitory population.
f_E	Firing rate functions for the excitatory population.
f_I	Firing rate functions for the inhibitory population.
W_{EI}	Connection weights between excitatory and inhibitory.
W_{IE}	Connection weights between inhibitory and excitatory.
W_{EE}	Self-excitation weights.
W_{II}	Self-inhibition weights.
r	Absolute refractory period.

A.1.2 Equations 2.3 & 2.4

I_{ION}	Ionic current.
I_K	Potassium current.
I_{Na}	Sodium current.
I_L	Leakage current.
C_M	Membrane capacitance.
V	Resting potential.
V_K, V_{Na} and V_L	Ionic channel potentials.
R_K, R_{Na} and R_L	Channel resistance.
V_m	Membrane potential.
I_{STIM}	Applied stimulation current.

Q_{tot}	Total charge injected.
T	Pulse width for a mono-polar stimulation.

A.2 Symbols and Notations Used in Chapter 3

A.2.1 Equations 3.1, 3.2 & 3.3

TP	True positives.
FP	False positives.
FN	False negatives.
TN	True negatives.
PD^+	PD periods.
PD^-	PD-free periods.
$Error_{weighted}$	Weighted classification error.
F_1	F1-score.
P	Precision.
R	Recall.
M_{CC}	Mathews Correlation Coefficient.

A.3 Symbols and Notations Used in Chapter 4

A.3.1 Equation 4.1

s_j	j-th independent source signals.
x_i	i-th discrete point in the original signal.
a_{ij}	Mixing matrix.

A.3.2 Equations 4.2, 4.3, 4.4, 4.5, 4.6 & 4.7

X_t	Stochastic process.
Z_t	Independent white noise error terms.
q	Order of MA Process.
β_i	Parameters of MA Process.
p	Order of AR process.
α_i	Parameters of AR process.
c	Constant term.
μ	Mean (average) of a series.
σ_Z^2	Variance of a white noise process.
τ	Translation parameter.
t	Time.
$E[X(t)]$	Expectation of the joint distribution $X(t)$.
$Cov[X(t), X(t+\tau)]$	Covariance of $X(t)$.
Δ	Differencing operator.
ε_t	Mean zero innovation process.
y_t	Time series.

A.3.3 Equations 4.8

$\gamma(k)$	Partial autocorrelation between y_t and y_{t+k}.
$Cor(y_{t+k}, y_t)$	Autocorrelation between y_t and y_{t+k}.
$P_{t,k}(y_t)$	Parameter adjusting for partial autocorrelation.

A.3.4 Equations 4.10

$Corrnorm_{x,y}$	Normalized cross-correlation.
$a[n]$ and $b[n]$	Two data sequences, where n is an integer.

A.3.5 Equation 4.11

r	PD rate.
Δt	An interval of time.
d	Number of PD events over an interval Δt.

A.3.6 Equation 4.12

L	Bit quantisation per sample.
$CompCost_a$	Complexity cost of the combination of PD detection algorithm labelled a.
NOP_a	NOP of the combination PD detection algorithm labelled a.
$area_a$	Area of the combination PD detection algorithm labelled a.
v	Number of combinations of PD algorithms.

A.3.7 Equations 4.13

$ErrCost_a$	Error cost of the combination of PD detection algorithm labelled a.
F_1	F1-score.
C_P	Choice Probability.

A.3.8 Equations 4.14

$X_n[t;f]$	STFT for the discrete time index t and discrete frequency index f.
$x[n]$	Discrete input signal with index n.
W	Window length.

A.3.9 Equations 4.15

$W(u, 2^j)$	DWT with u as the translational parameter and 2^j as the scale parameter.
Ψ	Wavelet function.

A.3.10 Equations 4.16, 4.17 & 4.18

A	Reduced feature vector.
B	Original feature vector.
U	Projection matrix.
PC	Square matrix containing the coefficients of all principal components.
m	Number of training examples.
n	Feature vector dimension.
H	Projection matrix for NMF.

A.3.11 Equations 4.19, 4.20, 4.21, 4.22, 4.23, 4.24, 4.25 & 4.26

$O = [o_1, o_2, \ldots, o_D]^T$	Vector of training examples having a total of D observations.
n	Feature vector dimension.
$o_i = [o_i^{(1)}, o_i^{(2)}, \ldots, o_i^{(n)}]$	i-th training example.
$C = [C_0, C_1, \ldots, C_k]$	Vector of $k+1$ classes, with each class having a total number observations $D_0, D_1, \ldots, D_k$ respectively.
μ_i and σ_i^2	Mean and variance for class C_i.
S_D	Interclass distance.
S_V	Intraclass variance.
$\mu_i = [\mu_i^{(1)}, \mu_i^{(2)}, \ldots, \mu_i^{(n)}]$	Mean for class i having n-dimensional features.
$\sigma_i = [\sigma_i^{(1)}, \sigma_i^{(2)}, \ldots, \sigma_i^{(n)}]$	Standard deviation for class i having n-dimensional features.
$J(S_D, S_D)$	Fisher's criterion of separability.

A.3.12 Equation 4.27

$f_{kNN}(\lambda)$	k-NN discriminating function.
λ	Test case.
Υ_n	Labels for the training datasets.
$N_k(\lambda)$	The index of k-nearest neighbours of test case, λ.

A.3.13 Equations 4.28 & 4.29

$f_{LR}(\lambda)$	LR discriminating function.
Υ	Class labels for LR (0 for non-PD or 1 for PD).
$\lambda^{(i)}$	Extracted features for test case i, with a label $\Upsilon^{(i)}$.
θ	Vector of threshold and weight parameters.

A.3.14 Equations 4.30

$f_{SVM}(\lambda)$	SVM discriminating function.
$K(\lambda_i, \lambda)$	Kernel transformation function.
λ_i	Support vectors with their labels Υ_i and weights ρ_i.
ω	Classification threshold.

A.3.15 Equations 4.31

$Training\ NOP$	Total NOP during training.
$Operating\ NOP$	Total NOP during normal operation.

A.4 Symbols and Notations Used in Chapter 5

A.4.1 Equation 5.1

P_i	Average power for level i coefficients.
N_i	Number of coefficients in sample i.
$x_i[n]$	DWT coefficients at level i.

A.4.2 Equations 5.2 & 5.3

h_0	Half-band low pass filter.
g_0	Half-band high-pass filter.
d_i	Detail coefficients at level i of DWT.
a_i	Approximation coefficients at level i of DWT.
$a_i(k)$	k is the index of the input signal (approximation coefficients, a_i).
$h_0(n)$	n represents the index of the filter coefficients of h_0.

A.5 Symbols and Notations Used in Chapter 6

A.5.1 Equation 6.1

$f_{GMM}(x)$	State Estimation function for GMM.
w_i	Weight assigned to a particular Gaussian model i.
$\vec{x}$	Input feature vector for test case.
μ_i and Λ_i	Mean and covariance vectors for Gaussian model i.
N	Number of Gaussian models.

A.5.2 Equation 6.2

$Error_{weighted}$	Weighted classification error.
TP	True positives.
FP	False positives.
FN	False negatives.
TN	True negatives.

A.5.3 Equation 6.3

NOP	Average number of operations.
$N_{add(sub)}$	Number of 1-bit additions or subtractions.
$N_{mult(div)}$	Number of 1-bit multiplications or divisions.
Res	Resolution of the data converter.

A.5.4 Equation 6.4

$\mu_A(a; x_i, y_i, z_i)$	μ_A is the certainty of membership in a fuzzy set. a is the universe of discourse. x_i,y_i and z_i are scalar parameters representing points on the triangle.

A.5.5 Equation 6.5

$\mu_D(d; \sigma_i, d_i)$	μ_D is the certainty of membership in a fuzzy set. d is the universe of discourse. σ_i and d_i are scalar parameters representing the standard deviation and mean respectively.

A.5.6 Equation 6.6

$\mu_D(d;\sigma_{ia},\sigma_{ib},d_i)$	μ_D is the certainty of membership in a fuzzy set. d is the universe of discourse. σ_{ia} is the standard deviation of the left tail of the Gaussian membership function. σ_{ib} is the standard deviation of the right tail of the Gaussian membership function.

A.5.7 Equation 6.7

C_{stim}	Incremental stimulation frequency.
A_i	The area under the membership function C_i.